An Introduction to Calculus and Algebra

VOLUME 1

Background to Calculus

Open University Set Book

The Open University Press

An Introduction to Calculus and Algebra

Volume 1 *Background to Calculus*

An Introduction to Calculus and Algebra

Volume 1

Background to Calculus

The Open University Press

The Open University Press
Walton Hall Bletchley
Buckinghamshire

First published 1971.
Second impression 1972

Printed in Great Britain by
The Staples Printing Group
at their Woking establishment

SBN 335 00002 9

1.2

Contents

NOTE

References to particular examples or exercises are made throughout by giving chapter, section, and example or exercise number; Example 4 in Chapter 1 section 1 would thus be referred to as Example 1.1.4, and so on.

Editors' Preface

This is the first of three volumes presenting some of the essential concepts of mathematics, a few important proofs (usually in outline), together with exercises designed to reinforce the understanding of the concepts and to develop the beginnings of technical skill.

The major part of the material used here has been selected from the correspondence texts of the *Open University Foundation Course in Mathematics*. Open University courses provide a method of study, at university level, for independent learners, through an integrated teaching system which includes textual material, radio and television programmes, local tutorial arrangements, and short residential courses. The correspondence text components of the Mathematics Foundation Course were produced by a Course Team (the names of the members are listed below) and were edited by Professor M. Bruckheimer and Dr. Joan Aldous.

The selection of material for these three volumes has been made with the needs of students of other subjects particularly in mind, by the Course Team preparing the second-year short course *Elementary Mathematics for Science and Technology*, and the three volumes constitute the set book around which the course is designed. (The members of this Course Team are listed below.)

In preparing these volumes, the Course Team has attempted to provide the kind of mathematics which is particularly useful for students who already have some knowledge of science or technology, but who, before proceeding in their own subjects, need to deepen their appreciation of the mathematical concepts underlying the techniques of calculus and algebra.

The special character of the original Foundation Course texts has been preserved as far as is possible, but the scope of these volumes is narrower than that of the course, which endeavours to give an overall picture of mathematics. For a much fuller appreciation of what mathematics is and what mathematics does, the reader is therefore referred to the original Mathematics Foundation Course correspondence texts.

WALTON 1971

GRAHAM FLEGG
ROGER MEETHAM

Mathematics Foundation Course Team

Professor M. Bruckheimer
Dr. J. M. Aldous
Mr. D. J. A. Burrows
Mr. R. Clamp
Mr. S. N. Collings
Dr. A. Crilly
Dr. D. A. Dubin
Mr. H. G. Flegg
Mr. E. Goldwyn
Mr. N. W. Gowar
Dr. A. Graham
Mr. R. D. Harrison
Mr. H. Hoggan
Mr. F. C. Holroyd
Miss V. King
Mr. R. J. Knight
Dr. J. H. Mason
Mr. R. Nelson
Miss J. Nunn
Professor R. M. Pengelly
Professor O. Penrose
Dr. G. A. Read
Mr. J. Richmond
Mr. E. Smith
Professor R. C. Smith

Course Assistants

Mr. J. E. Baker
Mr. W. D. Crowe

General Course Consultant

Mr. D. E. Mansfield

Elementary Mathematics for Science and Technology Course Team

Professor R. M. Pengelly
Mr. H. G. Flegg
Dr. A. R. Meetham
Mr. L. Aleeson
Mr. G. Burt
Dr. J. K. Cannell
Mr. R. Clamp
Dr. P. M. Clark
Mr. S. N. Collings
Dr. A. Cooper
Dr. A. Crilly
Professor M. J. L. Hussey
Mr. E. G. Law
Mr. F. B. Lovis
Mrs. V. Richards
Dr. R. A. Ross
Dr. T. B. Smith

Course Assistant

Mr. R. W. Duke

Notation

The symbols are presented in the order in which they appear in this volume.

Symbol	Meaning	Page
$\{a, b, c, d, \ldots\}$	The set of elements $a, b, c, d, \ldots$	5
$a \in A$	a is an element of the set A	5
Z	The set of all integers	6
Q	The set of all rationals	6
I	The set of all irrationals	6
R	The set of all real numbers	6
$\{x : x$ has some property $P\}$	The set of all elements x which have the given property P	6
$A \subset B$	The set A is a proper subset of the set B	8
$A \subseteq B$	The set A is a subset of the set B	8
$m: A \rightarrow B$	The mapping m maps the set A to the set B	12
$m: a \longmapsto b$	The image of a under the mapping m is b	12
$m(a)$	The image of a under the mapping m	12
$f = g$	The mappings f and g have the same domain and same rule	19
$x \leqslant a$	x is less than or equal to a	24
$x \geqslant a$	x is greater than or equal to a	24
$[a, b]$	The set of real numbers x such that $a \leqslant x \leqslant b$	24
$\lvert x \rvert$	The modulus of x	25
$P \times Q$	The Cartesian product of the sets P and Q	27
$\underset{\sim}{u}$	The sequence $u_1, u_2, u_3 \ldots$	48
$\lim \underset{\sim}{u}$	The limit of the sequence $\underset{\sim}{u}$	49
$f + g$	The sum of two functions	59
$f - g$	The difference of two functions	60
$f \times g$	The product of two functions	60
$f \div g$	The quotient of two functions	60
$g \circ f$	The function defined by $g \circ f: x \longmapsto g(f(x))$	63
$\lim\limits_{x \text{ large}} f(x)$	The limit of f for large numbers in its domain	84
$\lim\limits_{x \to a} g(x)$	The limit of g near the point a	86
$a \simeq b$	a is approximately equal to b	100
exp	The exponential function	104
e	$e = \exp(1) = 2.71828$	104
ln	The natural logarithm function	106
$\underset{\sim}{u} + \underset{\sim}{v}$	The sum of two sequences	123
$\int_a^b f$	The definite integral of f in $[a \,.\, b]$	150
Δh	The difference operator for the spacing h	177
f'	The derived function of f	180
D	The differentiation operator	186
I	The integration operator	225
$[F]_a^b$	$F(b) - F(a)$, that is $\int_a^b f$, where $DF = f$	231

CHAPTER 1 SETS AND MAPPINGS

1.0 Introduction

In this first chapter our main concern is to explain and define the terms *set* and *mapping*.

We start by looking at a number of very simple examples so that the two concepts are demonstrated before they are precisely defined.

Once the general idea of a set and a mapping is grasped, there is an immediate need for some appropriate notation. The notation which we shall use throughout is explained, and some further concepts are introduced so that precise definitions can be given.

The particular kind of mapping which we call a *function* is fundamental to mathematics and will appear in almost every chapter. This is defined in section 1.3, and will be later extensively discussed in Chapter 3 of this volume.

1.1 Groping for Definitions

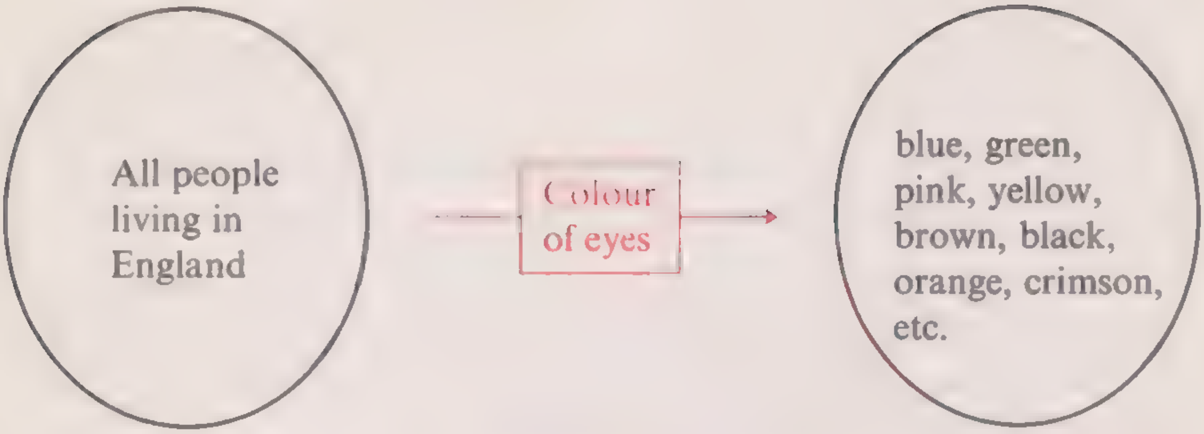
In this section we try to illustrate the ideas of *set* and *mapping* without making precise definitions.

Have a look at the following very simple examples each of which represents a mapping between sets.

Example 1

In this example a particular criminal corresponds (or we would say maps) to several of the fingerprints on record (since most people have more than one finger). Also all the fingerprints recorded in 1969 are only some of the fingerprints on record in 1971.

Example 2



If we take any person living in England we can use the rule Colour of eyes to associate one of the colours in the right-hand box to him. For example:

Fred Smith → $\boxed{\text{Colour of eyes}}$ → Blue

There are five things to notice about this example:

(i) We are told that the rule Colour of eyes applies to the set of all people living in England.

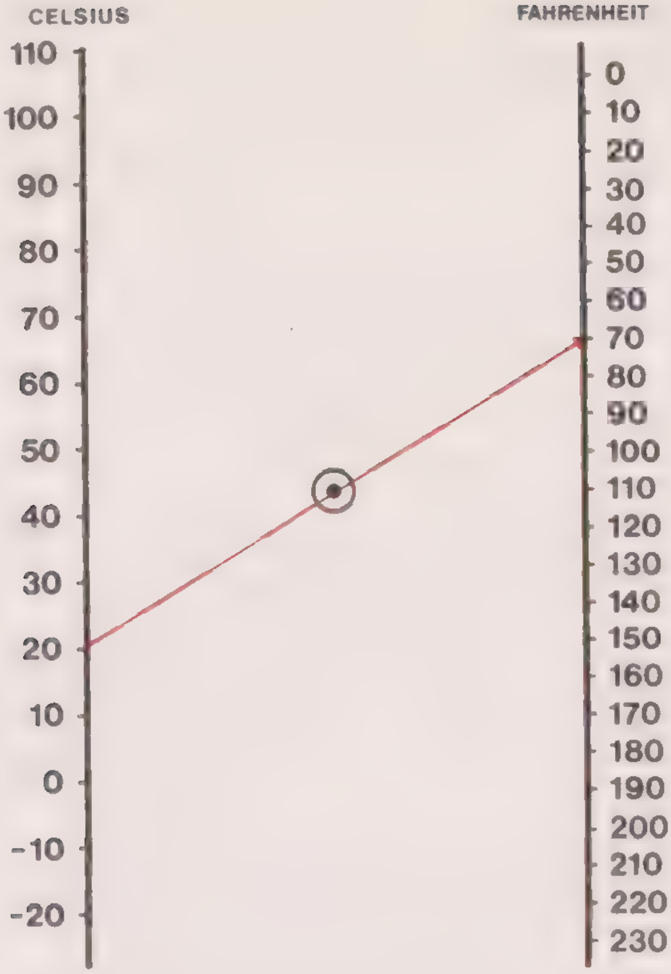
(ii) Each individual has a colour assigned to him by the rule (or possibly two colours, because there *are* rare individuals with eyes of different colours).

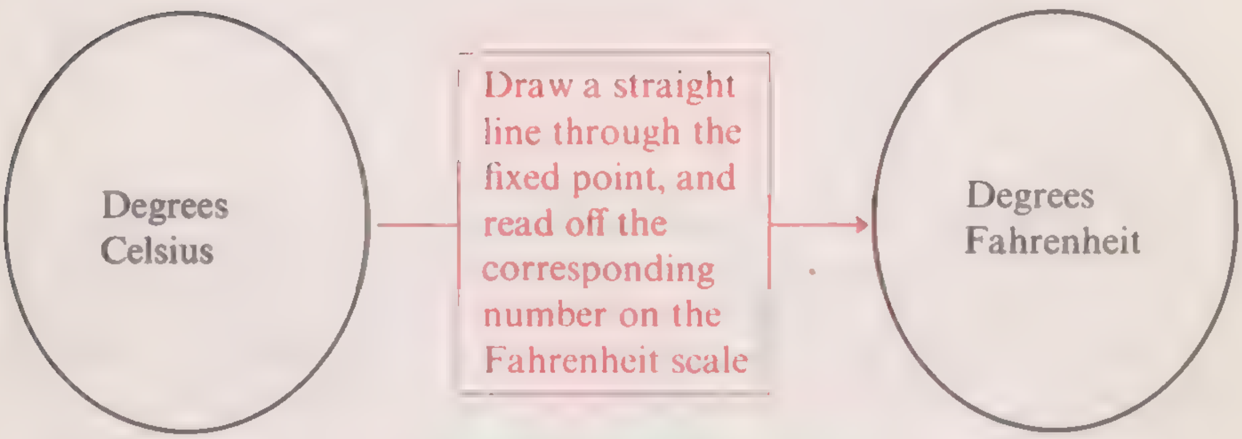
(iii) Many different people have the same colour assigned to them.

(iv) Not all of the colours on the right correspond to a person living in England. For example, crimson is listed even though no people have crimson eyes. You may have wondered why we included colours like crimson anyway. Strictly speaking we cannot be sure there are no crimson eyes *until* we have examined all the people.

(v) (There is a further point of a non-mathematical nature which ought to be mentioned. We are assuming that it is always easy to say exactly what colour the eyes are. In practice it may not be so simple. One person may say that your eyes are one colour, another person may say another. Some people even have eyes which appear to change colour in different lights.)

You may feel that the previous examples cannot really have anything to do with mathematics because we have not mentioned numbers. We are, of course, interested in numbers, but mathematicians are interested in a lot more besides.





Once again we have a rule Draw a straight line etc., which we are told applies to the set of numbers on the Celsius scale. For example:

$$0°\text{Celsius} \longrightarrow \boxed{\text{RULE}} \longrightarrow 32\,°\text{Fahrenheit}$$

and from the diagram we have

$$20°\text{Celsius} \longrightarrow \boxed{\text{RULE}} \longrightarrow 68\,°\text{Fahrenheit}$$

Try drawing some more straight lines on the diagram.

Notice particularly that in this example:

(i) Each temperature in °C corresponds to one, and only one, temperature in °F.
(ii) Each temperature in °F arises from one, and one only, temperature in °C.

The following situations are typical of many that arise in statistical surveys, scientific investigations and so on.

Example 4

The *Highway Code* gives the following table for stopping distances of a car travelling at various speeds.

Speed (mile/h)	20	30	40	50	60
Stopping Distance (ft)	40	75	120	175	240

Example 5

The following table gives the sales of electricity (in millions of kilowatt hours) for public lighting in Great Britain for the years 1938 to 1946.

Year	1938	1939	1940	1941	1942	1943	1944	1945	1946
Sale	367	248	17	18	20	20	28	177	260

Example 6

The following table was obtained when measuring the variation in temperature along a wire carrying an electric current. (The investigation was part of a research project aimed at developing a device for measuring electric current.) The distances are recorded in centimetres and measured from one end of the wire. The temperature is given in degrees Celsius.

Distance	2	4	6	8	10
Temperature	25	42	50	51	44

These last three examples all have a feature in common. In each case pairs of numbers are recorded. One number of the pair is taken from one set of numbers and another from a second set of numbers.

In all six examples two basic concepts are involved, that of a *set* of objects which may or may not be numbers and that of a *mapping* between sets of objects.

1.2 Sets

Mathematicians use the word set in a way which is very similar to its use in ordinary everyday speech. Roughly speaking, a set is a collection of objects. We do, however, need to be a little more precise and it is customary to define a set as a collection of distinct well-defined objects. This definition emphasizes two important properties of a set:

(i) no two objects belonging to a given set are identical,
(ii) for any object whatever, we can say whether or not it belongs to a given set.

So that we can refer more easily to particular sets we shall denote them by capital letters, such as A, B, . . ., X, Y, . . . For example, in Example 1.1.4 we have sets of numbers representing speeds and stopping distances. We could call these A and B, so that

A is the set $\{20, 30, 40, 50, 60\}$

and B is the set $\{40, 75, 120, 175, 240\}$.

Notice the way we write them: we list all the numbers of the set (*in any order*), separate them with commas, and enclose the whole lot with "curly brackets" (braces).

An object belonging to a set is called an element of the set (or sometimes a member of the set). We use a lower case letter to stand for an element, so we get statements like

"x is an element of X"

or just

"x belongs to X".

We use statements like this so often that we find it convenient to have a symbol in place of the words

"is an element of".

The symbol we use is $\in$, so that

"x is an element of X" becomes "$x \in X$"

but we still read it in the same way as before, or use the words "x belongs to X".

There are some sets of numbers that occur very frequently indeed in mathematics. For convenience we reserve certain capital letters to denote them, in particular the letters Z, Q, I, R. By combining these letters with "plus" and "minus" signs, we can represent twelve sets of numbers as follows:

Z^+	The set of positive integers
Z^-	The set of negative integers
Z	The set of positive and negative integers and zero
Q^+	The set of positive rationals
Q^-	The set of negative rationals
Q	The set of positive and negative rationals and zero
I^+	The set of positive irrationals
I^-	The set of negative irrationals
I	The set of positive and negative irrationals
R^+	The set of positive real numbers
R^-	The set of negative real numbers
R	The set of positive and negative real numbers and zero

(We shall use these letters to stand for these sets throughout.)

The set of all real numbers R is made up of the set of all rational numbers Q together with the set of all irrational numbers I.

The set of all rational numbers is made up of the set of all integers Z together with all real numbers that can be represented by fractions of the form m/n, where m and n are integers and m is not divisible by n.

Any real number that is not rational is irrational. The set of positive integers Z^+, i.e. $\{1, 2, 3, \ldots\}$, is often called the set of natural numbers.

We often want to refer to sets which have no conventionally accepted symbol to represent them. There is a standard way of writing such sets. A typical form of words is

"the set of all x such that x has some property"

This is conventionally written in mathematical shorthand as follows:

$\{x : x \text{ has some property}\}$

For example:

$\{x : x \text{ is a person with blue eyes}\}$

which we would read as:

> The set of *all* x such that x is a person with blue eyes

Notice that the colon in the notation corresponds to "such that" in the reading.
For sets of numbers, we have, for example:

$$\{x : x \in R \text{ and } x > 2\}$$

which we read as:

> The set of all x such that x belongs to the set of all real numbers and x is greater than 2

or, more briefly, as:

> The set of all real numbers greater than 2

Sometimes it is more convenient to refer to a particular set by listing all its elements, though, of course, we can only do this if the number of elements is finite. The order in which the elements are listed is immaterial since we are interested in the set as a whole, and so in Example 1.1.4 we could equally well write

A as $\{20, 50, 30, 60, 40\}$

instead of $\{20, 30, 40, 50, 60\}$.

Because of our definition requiring that each element of a set be *distinct*, we do not repeat any element in our list even if one or more elements appear more than once in a table such as that of Example 1.1.5. The set of numbers which gives the sale of electricity is

$$\{367, 248, 17, 18, 20, 28, 177, 260\}.$$

The number 20 appears only once in the list, even though it occurs twice in the table.

In the same way in Example 1.1.2 we could list the colours:

{Blue, Green, Pink, Yellow, Brown, Black, Orange, Crimson, etc.}

We list "Blue" once only, even though hundreds of people have blue eyes.

Equality of Sets

Two sets are said to be equal if they contain the same elements. Thus we write

$$\{1, 2, 3\} = \{3, 1, 2\}$$

or again,

$$\{a, b, c, d, e\} = \{c, e, b, d, a\}.$$

Subsets

Any set of elements chosen from a set is called a subset. For example:

(i) $\{367, 20, 260\}$ is a subset of $\{367, 248, 17, 18, 20, 28, 177, 260\}$,
(ii) $\{a, b\}$ is a subset of $\{a, b, c\}$.

Strangely enough we say that

$\{a, b, c\}$ is a subset of $\{a, b, c\}$

but if we wish to say "a subset which is not just the original set" we say a proper subset.

To make the point clear:

$\{a, b\}$ is a proper subset of $\{a, b, c\}$;

$\{a, b, c\}$ is a subset, but not a proper subset, of $\{a, b, c\}$

We frequently use a symbol to stand for "is a subset of". We write

$$A \subseteq B$$

to stand for "the set A is a subset of the set B". For "is a proper subset of", we write

$$A \subset B$$

to indicate that A is a subset of B and that A is not equal to B.

Sometimes we want to refer to a rather special set, the set having no elements which we call the empty set (or null set). We denote this set by $\varnothing$. For example, the set of persons living in England and having crimson eyes is equal to $\varnothing$. The empty set is a subset of every set.

Exercise 1

In each of the following questions, indicate which (if any) statements are correct.

(i) If $A = \{367, 20, 260\}$ and $B = \{367, 248, 17, 18, 20, 28, 177, 260\}$, then

(a) $A \subset B$
(b) $B \subseteq A$
(c) $A = B$
(d) $B \subset A$
(e) $A \subseteq B$

(f) A is a proper subset of B
(g) A is a subset of B

(ii) If $A = \{\text{Jim, Mary}\}$ and $B = \{\text{Blue, Green}\}$, then
(a) $A = B$
(b) $A \subset B$
(c) $B \subset A$

1.3 Mappings

In section 1.1 we referred to the idea of a mapping between sets. We have called the sets in Example 1.1.4 A and B. We can thus say that

To each number in A a number in B is assigned

or

The set A is mapped to the set B

or, in symbols,

$$A \longrightarrow B$$

which we read as "A maps to B".

We can now make a first attempt at a definition of the term "mapping":

> The essential feature of a MAPPING is that we have two sets and a method of assigning to *each* element of one set one or more elements of the other.

Example 1

We can map the set of all towns in the British Isles to the set of points on a piece of paper, by drawing a geographical map of the British Isles on the paper.

Example 2

We can map the set of all people to the set of all integers, using the rule that to each person is assigned his height in centimetres measured to the nearest centimetre.

(Notice that there are numbers in the second set in this example which are not assigned to any person in the first set. We don't know of anybody 2 cm. high.)

Example 3

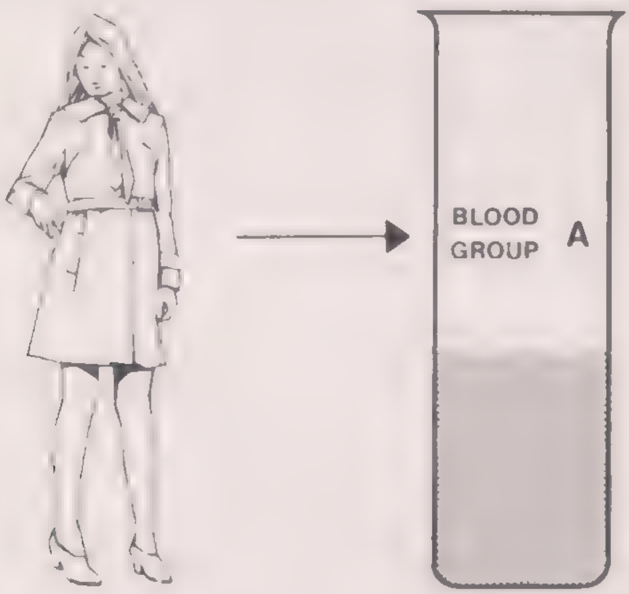
We can map the set of natural numbers to the set of natural numbers by assigning to each number its factors. For example:

6 maps to $\{1, 2, 3, 6\}$.

Example 4

We can map the set of all people to the set of all blood groups, by assigning to each person his blood group.

For example, Blood Group A might be assigned to a young lady called Elsie Tanner.



It is convenient to have a shorthand notation for statements like "Elsie Tanner maps to Blood Group A."

We already have the notation

$$A \longrightarrow B$$

which stands for "set A maps to set B".

What we need is a notation which says that a *particular* element of A has assigned to it a *particular* element or set of elements of B.

If a belongs to A (i.e. $a \in A$) and b belongs to B (i.e. $b \in B$), the statement "b is assigned to a" is abbreviated to

$$a \longmapsto b$$

Instead of the precise "b is assigned to a" we also often say "a maps to b", where the context makes clear which we mean.

Notice that we have just added a small bar to the arrow to indicate that this is a precise assignment of elements rather than just a mapping of one set to another, where there may be elements in the second set which are not assigned to elements in the first. For example we should have had a barred arrow in the illustration of Example 4, but it would be incorrect to put a barred arrow in the statement $A \longrightarrow B$ unless we know that *every* element of B is assigned to an element of A. So in Example 2, $A \longmapsto B$ is incorrect, but in Example 4, $A \longmapsto B$ is correct.

Images

if $a \in A$, $b \in B$

and $a \longmapsto b$,

we say that b is the IMAGE of a.

For example, if

Elsie Tanner $\longmapsto$ Blood Group A

then

Blood Group A is the image of Elsie Tanner.

This image of an element may be a set of elements, as we saw in Example 3, where a natural number has assigned to it its factors, e.g.

$$6 \longmapsto \{1, 2, 3, 6\}.$$

We could consider an image to be a set even if it consisted of only one element, but we do not make the distinction in this context between an element and the set containing that single element. For example, in the above we write $\{1, 2, 3, 6\}$ because when the image consists of more than one element it is convenient to think of it as a set. But although there is a logical distinction between Group A and {Group A}, for example, we shall ignore this distinction.

Naming a Mapping

If we wish to refer to a mapping, it is not always convenient to have to describe it in full every time, so we often use a symbol to represent it.

For example, the mapping of the set of people to the set of blood groups could be represented by the letter h. We then write expressions like

$$h: \text{Set of people} \longrightarrow \text{Set of blood groups}$$

involving whole sets, and

$$h: \text{Elsie Tanner} \longmapsto \text{Group A}$$

involving elements of sets.

Using the letter h we also write h (Elsie Tanner) to mean the image of Elsie Tanner under the mapping h. Therefore, we have:

$$h(\text{Elsie Tanner}) = \text{Group A.}$$

Either of the last two statements in red can be read "h maps Elsie Tanner to Group A" or "the image of Elsie Tanner under the mapping h is Group A", or "the mapping h, such that Elsie Tanner maps to Group A". Any other convenient letter may, of course, be used in place of h provided that we define what it means.

Summary of Notation

$a \in A$	means	a is an element of set A.
$m: A \longrightarrow B$	means	The mapping m maps the set A to the set B.
$m: a \longmapsto b$	means	The mapping m maps the element a to the element b, OR b is the image of a under (the mapping) m.

If the need arises we extend the notation still further and write, for example,

$$h: \text{set of all people} \longmapsto \text{set of all blood groups}$$

when every element of the image set corresponds to at least one element of the first set.

Discussion of Definitions and Notation

The definitions and notation that we give are by no means the only ones possible. There is far more freedom in mathematics than is commonly thought. We choose our particular definitions because we consider them to be a workable set of definitions at our present level of discussion, and we have included only those which we need for our immediate purposes.

Other people may use different definitions, notation and terminology, and you should be careful to check them in any book you read.

Choices like these arise as frequently in mathematics as in any other subject. In mathematics, however, we consider it important that once we have decided on a set of rules, we have got to play our game strictly in accordance with them; that is, until we change the rules, and declare new ones.

Example 5

We reproduce the tables of Examples 1.1.4, 1.1.5 and 1.1.6 here for your convenience.

Table I

s

Speed (mile/h)	20	30	40	50	60
Distance (ft)	40	75	120	175	240

Table II

f

Year	1938	1939	1940	1941	1942	1943	1944	1945	1946
Sale	367	248	17	18	20	20	28	177	260

Table III

t

Distance	2	4	6	8	10
Temperature	25	42	50	51	44

The image of 20 under s is 40 and so we write

$$s: 20 \longmapsto 40$$

or

$$s(20) = 40$$

The image of 1943 under f is 20 and so we write

$$f: 1943 \longmapsto 20$$

or

$$f(1943) = 20$$

The image of $\{2, 4, 6\}$ under t is $\{25, 42, 50\}$ and so we write

$$t: \{2, 4, 6\} \longmapsto \{25, 42, 50\}$$

or

$$t(\{2, 4, 6\}) = \{25, 42, 50\}$$

If $A = \{2, 4, 6, 8, 10\}$ and $B = \{25, 42, 50, 51, 44\}$ we can write

$$t: A \longmapsto B \quad \text{and} \quad t(A) = B$$

But if $C = \{25, 42, 50, 51, 44, 99\}$, we *cannot* write $t: A \longmapsto C$ or $t(A) = C$ because 99 has no corresponding element in A, but we *can* write

$$t: A \longrightarrow C \quad \text{or} \quad t(A) \subseteq C$$

Exercise 1

Complete each of the following (i.e. replace the question mark appropriately where s, f and t are the mappings in Example 5).

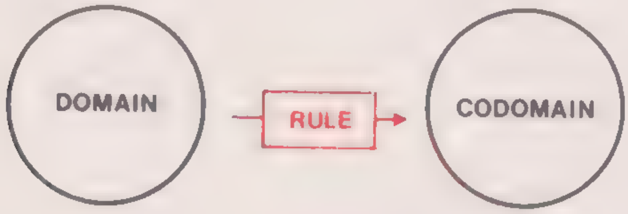
(i) The image of 30 under s is?
(ii) $t: 8 \longmapsto ?$
(iii) $f: ? \longmapsto 28$
(iv) $f(1942) = ?$
(v) $s(60) = ?$
(vi) $t(?) = 44$
(vii) $s: \{30, 40, 50\} \longmapsto ?$

Domain and Codomain

There are several points still to tidy up if we wish to be more precise about the meaning of words such as MAPPING.

> If we have a mapping from a set A to a set B, then we call A the DOMAIN of the mapping, and B the CODOMAIN of the mapping.

> The mapping involves some method, or rule, whereby to **each** element of the domain an image is assigned, and the codomain contains all the images.



In Example 1.1.2 (the colour of eyes example) the domain is:

the set of all people living in England.

The codomain is:

{Blue, Green, Pink, Yellow, Brown, Black, Orange, Crimson, etc.}

and, for example,

the image of Fred Smith is Blue

In our first tentative definition of a mapping earlier we required that to each element of the domain we must be able to assign an element (or elements) of the codomain. In other words each element of the domain must have an image in the codomain.

On the other hand, as we have seen in Example 1.1.2, there is no reason why the codomain should not include elements which are not images, for there is no question of applying a formula, or a rule, to elements of the codomain. Indeed, it is convenient not to insist that every element in the codomain be an image of an element in the domain. How, for example, could we predict the exact set of intelligence quotients (the codomain) of a set of people (the domain) before measuring them? All we know is that all the intelligence quotients will be found in the set of natural numbers, and so we could choose this set as the codomain.

All we require of the codomain is that it should contain the set of all images of elements in the domain. Thus, for instance, in the "colour of eyes" example we do no harm to include crimson as one of our colours, even though it is not the colour of anyone's eyes.

We are now in a position to make some definitions. *Notice that in the first definition we define a mapping to consist of three things (the domain, the rule, and a codomain).*

Summary of Definitions

> A MAPPING consists of a set A, a set B and a rule by which an element (or set of elements) of B is assigned to *each* element of A.

> The set A is the DOMAIN of the mapping.

The set B is the CODOMAIN of the mapping.

If an element a of the domain has assigned to it an element b or a set of elements T of the codomain, then b or T is the IMAGE of a. Each element of T is called *an* image of a. If T contains only one element, b, then b is *the* image of a.

If T is the set of all elements of the codomain which are images of elements of the subset S of the domain, then T is the IMAGE of S.

The mappings for which each element in the domain has a unique element as its image are particularly important, and so we give them a special name.

A FUNCTION is a mapping for which each element in the domain has only one element as its image.

Note

It is becoming increasingly common usage to use the terms "mapping" and "function" synonymously. What we have called a mapping is sometimes referred to as a "correspondence". We have kept to our definition of mapping because we feel that the word mapping carries with it more of a sense of movement from one set to another than the word correspondence. As with any mathematical term, you must be sure, if you consult a text book, of the way in which the author is using the term.

Exercise 2

In each case state whether the statement is true or false:

(i) A mapping is always a function.
(ii) The domain of a mapping is the set of all images under the mapping.
(iii) If m is a function with domain A and codomain B, then $m(a)$ must be an element of B. $(a \in A)$
(iv) The set of all images under a mapping is called the

codomain of the mapping.

(v) If $A = \{\alpha, \beta, \gamma\}$ and $B = \{1, 2, 3\}$ then the list:

$$m : \alpha \longmapsto \{1, 2\}$$

$$m : \beta \longmapsto 1$$

$$m : \gamma \longmapsto 1$$

defines a mapping from A to B.

(vi) The list in (v) defines a function.

(vii) If $A = \{\alpha, \beta, \gamma\}$ and $B = \{1, 2, 3\}$ then the list:

$$m : \alpha \longmapsto \{1, 2\}$$

$$m : \beta \longmapsto 3$$

defines a mapping from A to B.

Rules for Mapping Numbers to Numbers

When the domain and codomain of a mapping are sets of numbers, we can often abbreviate the rule which tells us how to find the image of any element of the domain by using common algebraic notation.

Example 6

The mapping with domain R (the set of all real numbers); codomain R, and rule

$$\boxed{\text{Double it}}$$

could be described by saying

For any element x in the domain $x \longmapsto 2x$

Example 7

Let f be the mapping with domain and codomain R and the rule

for any real number	(i)	square it
	(ii)	multiply the result by 6
	(iii)	subtract from this result twice the original number
and then	(iv)	add 1

We usually abbreviate this and refer to the mapping

$$f: x \longmapsto 6x^2 - 2x + 1 \quad (x \in R)$$

Any letter could have been used in place of x as an arbitrary element of the domain; however when the domain is R it is common practice to use x. Any letter used in this way, to make a statement about every element of a set, is called a VARIABLE.

The statement $x \in R$ in the formula above, contains rather a lot of information. It tells us that

(i) the domain is R

and

(ii) the letter x is a variable, which can take any value in the domain

Example 8

The statements

$$f: x \longmapsto 2x^2 - 3 \quad (x \in R)$$
$$f: t \longmapsto 2t^2 - 3 \quad (t \in R)$$
$$f: a \longmapsto 2a^2 - 3 \quad (a \in R)$$

are all equivalent, and all define the same function; each of the statements

$x \in R$

$t \in R$

$a \in R$

specifies the domain (as the set R), and each declares the corresponding letter to be a variable. On the other hand, the statements

$$f: x \longmapsto 2x^2 - 3$$

and

$$f: 6 \longmapsto 2 \times 6^2 - 3 = 69$$

do not define a mapping, for there is no mention of the domain. They are simply statements about what happens to *particular* elements under the mapping, in the first case the element x, in the second the element 6.

Our definition of a mapping states that we must (among other things) specify the codomain, and yet we seem to be ignoring it. Is it true that

$$f: x \longmapsto x^2 + 4x - 1 \quad (x \in R)$$

defines a mapping?

Strictly speaking the answer is NO, but if we assume that the codomain is also R then the definition is complete. Since, in general, given the domain and the rule we can work out the set of images (and the codomain can be any set containing the set of images), we shall not include the specification of the codomain in the definition of the mapping unless we have a particular interest in it.

Strictly speaking mappings are equal only if they have the same DOMAIN, CODOMAIN and RULE. However, we shall be content with having the same domain and rule. For example, in the mapping above

$$f: x \longmapsto x^2 + 4x - 1 \quad (x \in R)$$

we would not distinguish between the two cases where the codomain is R and where it is the set of real numbers greater than or equal to -5, which is the set of images. If f and g are two mappings with the same domain and rule, we shall write

$$f = g$$

Exercise 3

For each of the functions f, g and h defined below, state

(a) the domain,
(b) the image of the set $\{1, 2, 3\}$,
(c) the image of the domain.

(i) $f: x \longmapsto 2x + 1$ ($x \in$ the set of positive real numbers, which is denoted by R^+),
(ii) $g: x \longmapsto x^2 - 2$ ($x \in R$),
(iii) $h: x \longmapsto 3$ ($x \in R$).

1.4 Graphs

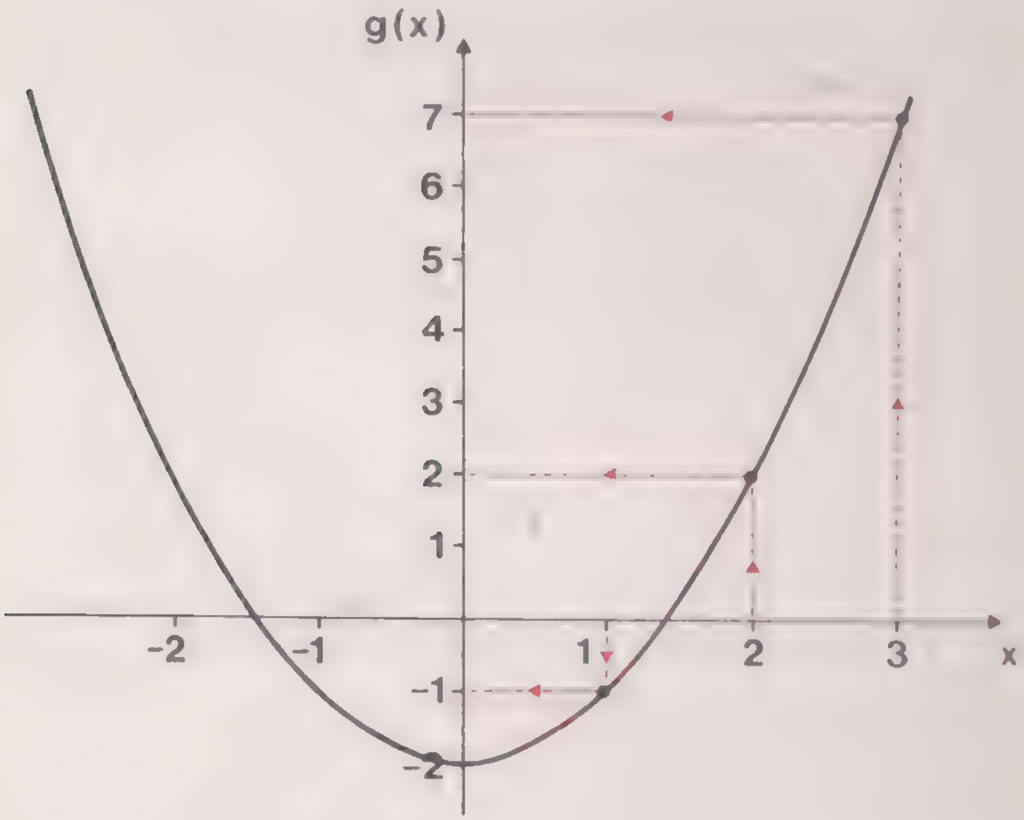
In this section we relate the terms we have introduced so far to graphs, with which you may be familiar.

Example 1

The function defined by

$$g: x \longmapsto x^2 - 2 \quad (x \in R)$$

has a graph which looks like this:



Graph of g

The mapping from the domain to the codomain can be visualized by following arrows drawn parallel to the vertical and horizontal axes

$$g: 1 \longmapsto -1$$

$$g: 2 \longmapsto 2$$

$$g: 3 \longmapsto 7$$

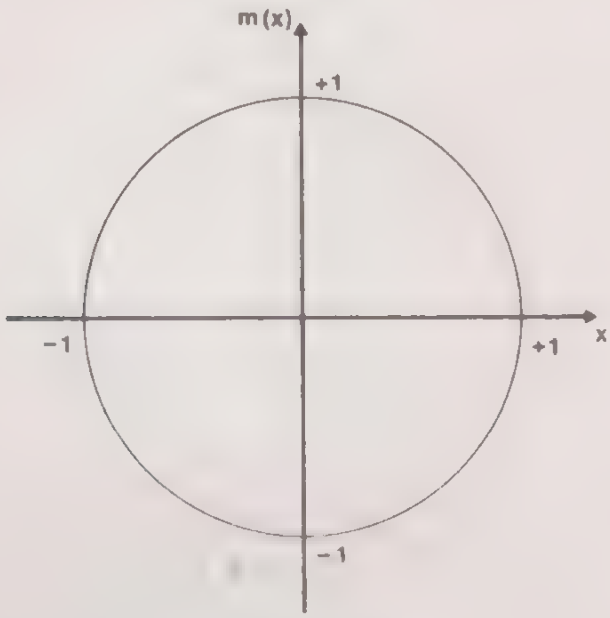
Many mappings which are not functions can also be represented graphically.

Example 2

If $[-1, 1]$ is the set of real numbers between -1 and $+1$ inclusive, the mapping m defined by

$$m: x \longmapsto \{\sqrt{1-x^2}, -\sqrt{1-x^2}\} \quad (x \in [-1, 1])$$

is certainly not a function, because each element of the domain (except ± 1) has two images, but it has a graph which looks like this:



Graph of m

N.B.

In some older mathematical textbooks, the term function is used to cover expressions equivalent to $x \longmapsto \{\sqrt{1-x^2}, -\sqrt{1-x^2}\}$ $(x \in [-1, 1])$. This does not conform to the definitions adopted in these volumes.

Exercise 1

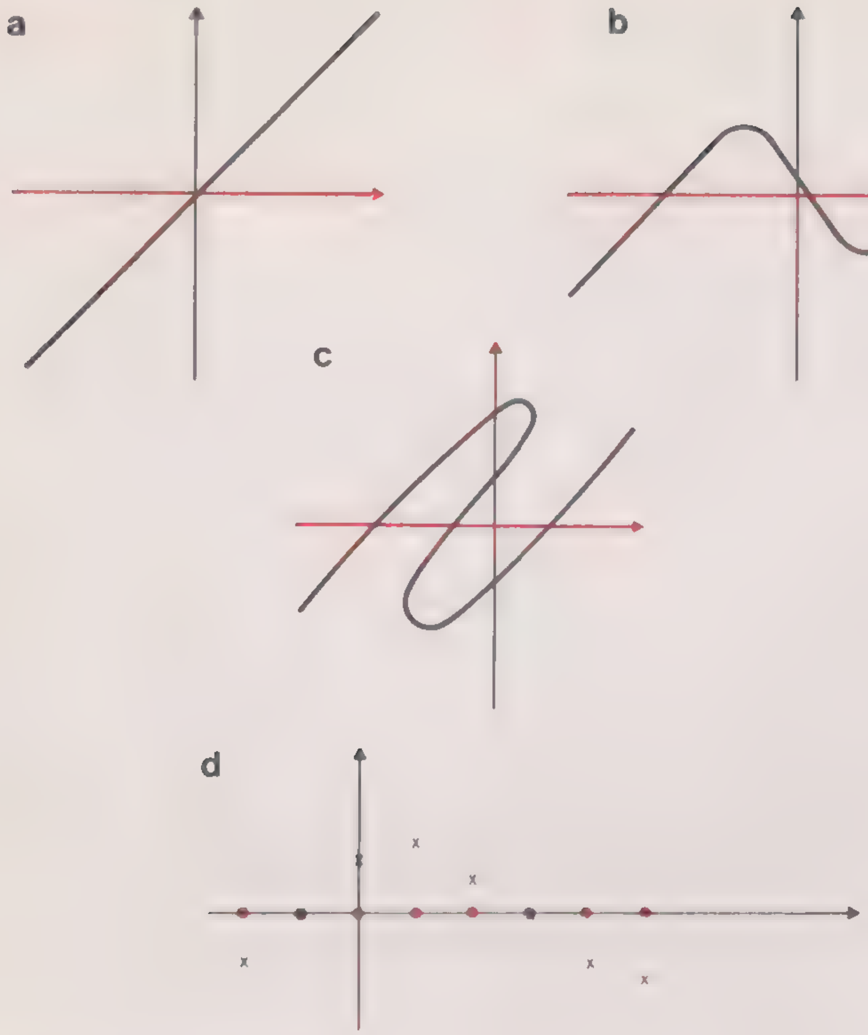
Indicate which of the following graphs are

(i) graphs of functions

or

(ii) graphs of mappings that are not functions.

By convention, the horizontal axis is used to show the domain and the vertical axis is used to show the codomain. The domain is shown in red.



Cartesian Co-ordinates

As we have said before, it is conventional when drawing graphs of mappings, whose domains and codomains are sets of numbers, to represent the domain on a horizontal axis and the codomain on a vertical axis, and we shall use this convention. We have seen that it is common to use a letter x to stand for a variable in the domain and for this reason the horizontal axis is usually called the x-axis.

Similarly the letter y is often used as a variable in the codomain and so the vertical axis is called the y-axis.

Exercise 2

Draw graphs for each of the following functions, and indicate the domain, the codomain and the image of the set $\{1, 2, 3\}$.

(i) $f : x \longmapsto 2x + 1 \quad (x \in R^+)$
(ii) $g : x \longmapsto x^2 - 2 \quad (x \in R)$
(iii) $h : x \longmapsto 3 \quad (x \in R)$

What Do We Mean by a Graph?

We have taken a GRAPH to mean a collection of points on a piece of paper; but unfortunately it is not always possible to represent mappings pictorially (in particular, if the domain and codomain are not sets of real numbers).

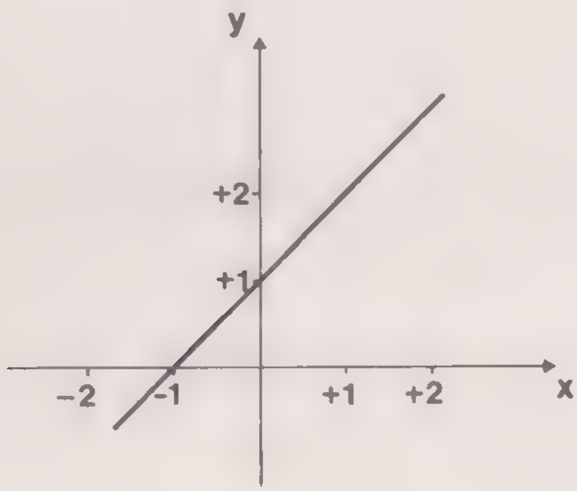
We might mention now that to overcome these problems some people define the graph of the mapping f to be the set of all pairs of elements (x, y), where $x \in A$, the domain of f, and y is the image of x under f (or an image, where the image is a set).

Using a Graph to Define a Function or Mapping

Basically, we use a mapping to tell us how to draw a graph. Sometimes we can reverse the process and use a graph to help define a mapping.

Example 3

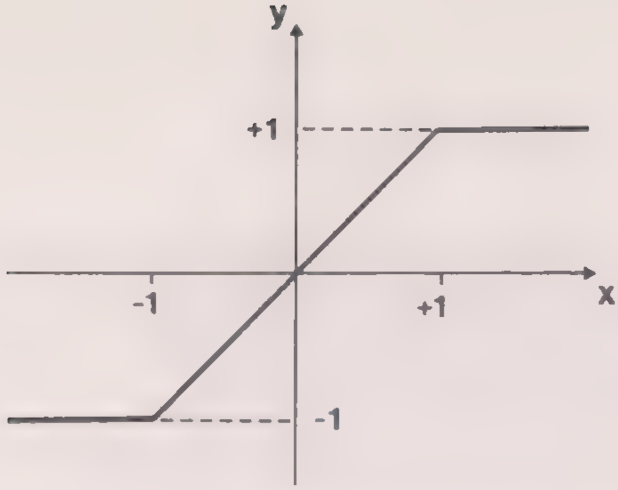
The graph



has the equation $y = x + 1$, and if we specify that x can take any real value in this equation then we have defined the function

$$f: x \longmapsto x + 1 \quad (x \in R)$$

Example 4



This graph, together with a statement that x can take any real value, say, defines a function. But can we express this function by a formula? The formula is

$$y = -1 \quad \text{when } x \leqslant -1$$
$$y = +1 \quad \text{when } x \geqslant 1$$

and

$$y = x \qquad \text{when } x \in [-1, +1]$$

N.B.

(i) $x \leqslant a$ means "x is less than or equal to a"

(ii) $x \geqslant b$ means "x is greater than or equal to b"

(iii) $x \in [a, b]$ means "x belongs to the set of all real numbers between a and b, including a and b".

(iv) Expressions containing the first two symbols (or $>$ or $<$) are known as inequalities.

Thus

$$\boxed{\begin{aligned} f(x) &= -1 \quad \text{if } x \leqslant -1 \\ f(x) &= +1 \quad \text{if } x \geqslant 1 \\ &\text{and} \\ f(x) &= x \qquad \text{if } x \in [-1, +1] \end{aligned}}$$

is the rule which defines the function f. The domain of f is R.

We could also write

$$f:x \longmapsto \begin{cases} -1 & \text{if } x \leqslant -1 \\ +1 & \text{if } x \geqslant +1 \\ x & \text{if } -1 \leqslant x \leqslant 1 \end{cases} \quad (x \in R)$$

Exercise 3

(i) Sketch the graph of the function f, where

$$f:x \longmapsto 1 + x \quad (x \in [-1, +1])$$

(ii) Sketch the graph of the function g, where

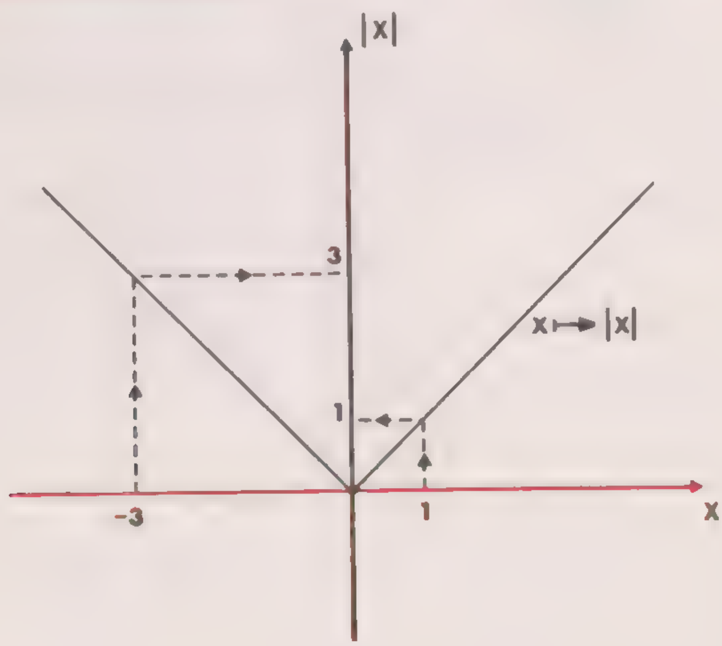
$$g:x \longmapsto 1 - x \quad (x \in [-1, +1])$$

The MODULUS Function

The function

$$f:x \longmapsto \begin{cases} x & \text{if } x \geqslant 0 \\ -x & \text{if } x < 0 \end{cases} \quad (x \in R)$$

occurs very frequently in mathematics and so we give it a name and we have a special notation for it. We denote $f(x)$ by the symbol $|x|$ and call it the modulus of x. Thus, for example, $|1| = 1$ and $|-3| = 3$. The function is called the modulus function.



1.5 Cartesian Products

One very precise way of specifying a mapping is to write down a list of ordered pairs. This is fine, but it is obviously tedious or even impossible unless the list contains only a few elements. Indeed, we have already seen this in Example 1.1.4. We could represent this mapping by the set of pairs of numbers

$$\{(20, 40), (30, 75), (40, 120), (50, 175), (60, 240)\}$$

Each pair consists of an element of the domain followed by its corresponding image in the codomain. Notice that the order in which the elements of any pair are written is important. Thus (20, 40) would not mean the same as (40, 20).

When the order in which the elements of a pair are written is important, we call the pair an ordered pair.

Example 1

Look at the mapping m from

$$A = \{a, b, c\}$$

to

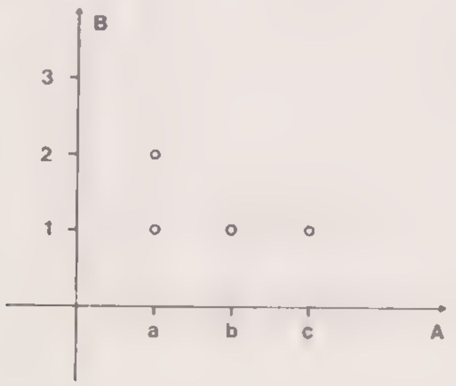
$$B = \{1, 2, 3\}$$

illustrated by

We can just as easily represent the mapping by the set of ordered pairs

$$\{(a, 1), (a, 2), (b, 1), (c, 1)\}$$

or by the diagram



Any mapping from A to B corresponds to a set of ordered pairs, and such a set must be a subset of the set of all possible ordered pairs,

$$\{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3), (c, 1), (c, 2), (c, 3)\}$$

However, our definition of mapping is such that not every subset of the set of all possible ordered pairs will define a mapping.

Exercise 1

Which of the following sets define mappings from the set A to the set B, where

$$A = \{a, b, c\}$$

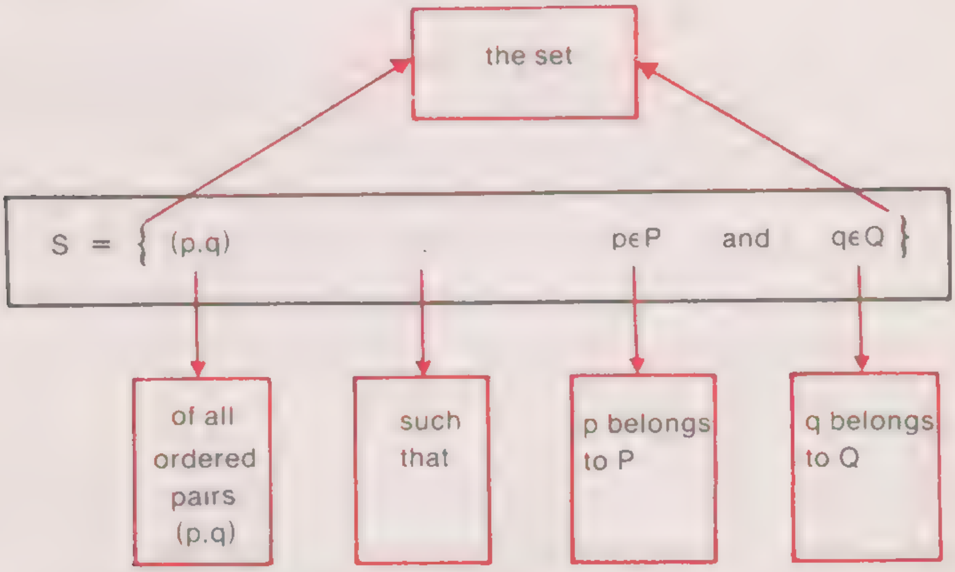
and

$$B = \{1, 2, 3\}$$

(i) $\{(a, 1), (b, 2), (c, 3)\}$
(ii) $\{(a, 1), (a, 2), (a, 3)\}$;
(iii) $\{(a, 1), (a, 3), (b, 2), (b, 1), (c, 3)\}$;
(iv) $\{(a, 1), (b, 1), (c, 2)\}$.

Which of these mappings are functions?

Any mapping from a set P to a set Q will correspond to a set of ordered pairs. The first element of each pair will belong to P, and the second element to Q. The set of ordered pairs will be a subset of the set of all possible pairs (p, q) where $p \in P$ and $q \in Q$. If we call this *set of all possible pairs S*, we have



The set S is called the Cartesian product of P and Q and we denote it by

$$P \times Q$$

(We read this as "*P cross Q*".)

The word "Cartesian" is derived from the surname Descartes. René Descartes was a famous seventeenth century French mathematician and philosopher who was the founder of *analytical geometry*, the application of algebra to geometry. His name was translated into Latin as "Renatus Cartesius", hence the adjective "Cartesian" which is perhaps best known when it appears in the phrase

"rectangular Cartesian co-ordinates"

describing a pair of axes used when drawing graphs in a plane.

Example 2

Let $P = \{\mathbf{K}, \mathbf{Q}, \mathbf{J}\}$

and

$$Q = \{\heartsuit, \clubsuit, \diamondsuit, \spadesuit\}$$

Then

$$P \times Q = \{(\mathrm{K},\heartsuit), (\mathrm{Q},\heartsuit), (\mathrm{J},\heartsuit), (\mathrm{K},\clubsuit), (\mathrm{Q},\clubsuit), (\mathrm{J},\clubsuit),$$
$$(\mathrm{K},\diamondsuit), (\mathrm{Q},\diamondsuit), (\mathrm{J},\diamondsuit), (\mathrm{K},\spadesuit), (\mathrm{Q},\spadesuit), (\mathrm{J},\spadesuit)\}$$

We can now consider mapping and function from a slightly different stand point as follows:

> A subset of $P \times Q$ defines a mapping from P to Q if every element of P appears as the first term of at least one ordered pair of the subset.

> A subset of $P \times Q$ defines a function from P to Q if each element of P appears as the first term of an ordered pair of the subset once and once only.

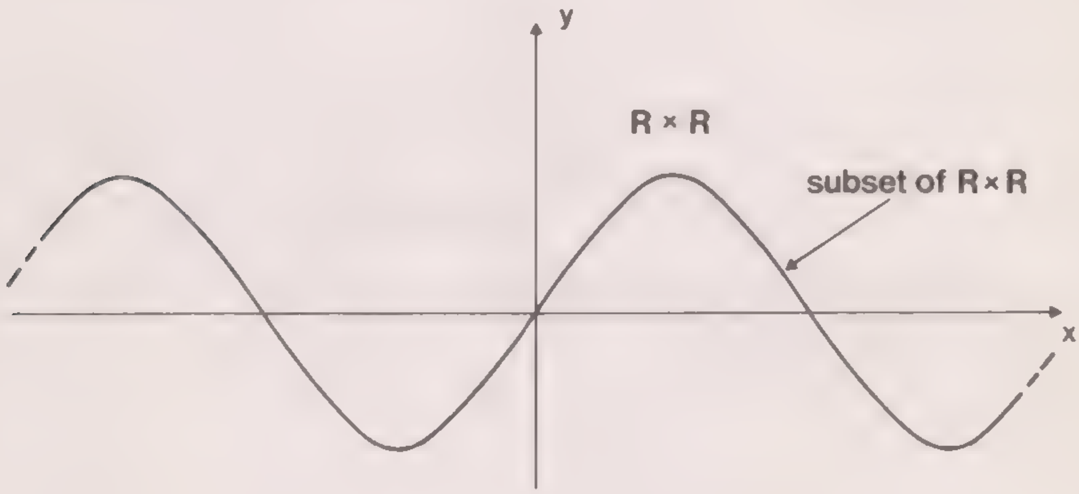
(Note that the codomain Q is not necessarily defined by the subset of $P \times Q$ in either case, because there is not a requirement that every element of the codomain should appear as the second term of at least one ordered pair of the subset. Each element of the domain and its image is specified, however, and so the set of images of the elements of P is defined, and it is this set which is of importance.)

Example 3

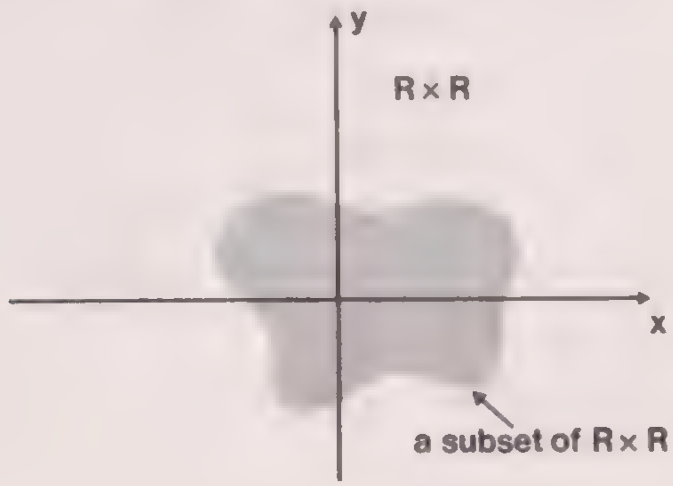
Probably the most frequently occurring collection of functions is the set of functions whose domain and codomain are subsets of the set of real numbers, R.

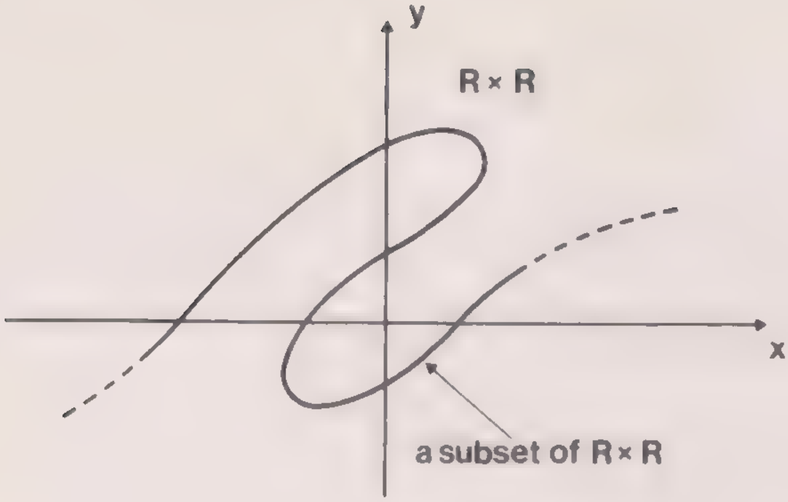
Each function of this kind can be identified with some particular subset of $R \times R$. Using the usual Cartesian co-ordinate system which we use for drawing graphs, the set $R \times R$ can be represented by the set of all points in a plane. Any subset of $R \times R$ then corresponds to a set of points in the plane.

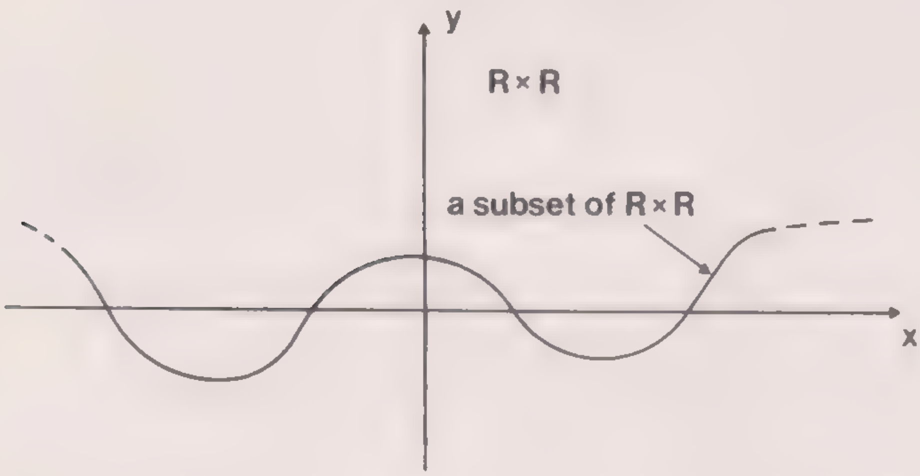
Thus the subset of $R \times R$ with which we identify, for example, the function $x \longmapsto \sin x$, is represented by



Some other possible subsets of $R \times R$ are represented by:







(The dotted ends of the curves are intended to mean that the curves extend indefinitely in the final direction indicated.)

Exercise 2

Why do *neither* of the first two of the three subsets of $R \times R$ depicted immediately above define a function $f: R \longrightarrow R$?

1.6 Equations and Inequalities

You are probably very familiar with the idea of solving an equation. But just what do we really mean when we say, for example,

"Solve the equation $f(x) = 0$"?

Let f be a mapping with domain, A, and codomain, B, which are sets of real numbers.

The phrase solve the equation $f(x) = 0$ means

find the set of all elements of A which map to 0 under f

We can write this set as

$$\{x : x \in A, f(x) = 0\}$$

and we call it the solution set of the equation $f(x) = 0$

Example 1

The solution set of the equation

$$x^2 - 1 = 0 \quad (x \in R)$$

is the set

$$\{-1, 1\}.$$

We thus have

$$\{x : x \in R, x^2 - 1 = 0\} = \{-1, 1\}.$$

Example 2

The solution set of the equation

$$x^2 + 1 = 0 \quad (x \in R)$$

is the empty set $\varnothing$.

Similarly, solve the inequality $f(x) > 10$, for example, means find the set of all elements of A which map to numbers greater than 10 under f.

We can write this set as

$$\{x : x \in A, f(x) > 10\}$$

and we call it the solution set of the inequality $f(x) > 10$.

Exercise 1

Illustrate on a number line the following sets. For example, the answer to (i) is

(i) $\{1, 1.5, 3.75\}$
(ii) $\{x : x \in R, 3 + x = 2\}$
(iii) $\{x : x \in R, x^2 + 3x + 2 = 0\}$
(iv) $\{x : x \in R, x > 0\}$
(v) $\{x : x \in R, x \leqslant 2\}$

Exercise 2

(i) Find a positive integer N such that

$$\frac{1}{N^2 + N + 1} < \frac{1}{10}$$

(ii) Find a positive integer N such that

$$\frac{1}{n + 1} < 0.01 \text{ for ALL integers } n \text{ greater than } N$$

That is, find, by determining a value for N, a set $A = \{n : n > N\}$ such that all elements of A belong to the solution set of the inequality (or, more briefly, such that A is a subset of the solution set).

(iii) Find a real positive number δ such that: $\delta \sin \delta < 0.1$

1.7 Additional Exercises

Exercise 1

(Refer to the list of symbols for special sets of numbers on page 6). State which of the following are true and which false.

(i) $R \subset Z$
(ii) $Z \subseteq R$
(iii) $Z \subset Z^+$
(iv) $Z^- \subset Q^-$
(v) $Q \subseteq I$
(vi) $I \subseteq Q$
(vii) $Z^+ \subseteq Q^+ \subseteq R^+$
(viii) $Z \subset I \subset R$

Exercise 2

If m is a mapping which maps a set A to a set B, and if $a \in A$ and $b \in B$, which of the following statements are true and which false?

(i) If $m : a \longmapsto b$, then $m(a) = b$
(ii) If $m : a \longmapsto b$, then $m(a)$ is the image of b
(iii) If $m : a \longmapsto b$, then $m(a) \in B$

Exercise 3

Say which of the following statements are true and which are false:

(i) The statement $f : x \longmapsto x^2 + 1$ $(x \in R)$ implies that $f(2) = 5$

(ii) The statement $f:x \longmapsto x^2 + 1$ implies that $f(1) = 2$ (Be careful!)

(iii) The statement $f(2) = 5$ implies that $f:x \longmapsto x^2 + 1\ (x \in R)$

(iv) The statement $f:x \longmapsto 2x + 6\ (x \in R^+)$ implies that $f(-10) = -14$

(v) The statement $f:x \longmapsto 2x + 6\ (x \in R)$ implies that $f:t \longmapsto 2t + 6$ where t is any real number

(vi) The statements $f:x \longmapsto 6x - 1\ (x \in R)$ and $f:t \longmapsto 6t - 1\ (t \in R)$ are equivalent

(vii) The statements $f:x \longmapsto 4x^2 + 1\ (x \in R)$ and $f:t \longmapsto 4t^2 + 1\ (t \in R^+)$ are equivalent

Exercise 4

(i) Sketch the graph of the function

$$f:x \longmapsto \begin{Bmatrix} 1 + x & \text{if } x \geqslant 0 \\ 1 - x & \text{if } x \leqslant 0 \end{Bmatrix}$$

with domain $[-1, +1]$

(ii) Sketch the graph of the function

$$g:x \longmapsto \begin{Bmatrix} |x| \text{ if } -1 \leqslant x < 1 \\ |2 - |x|| \text{ if } x < -1 \text{ or } x \geqslant 1 \end{Bmatrix} \quad (x \in R)$$

(iii) Sketch the graph of the function

$$h:x \longmapsto \begin{Bmatrix} \left|\frac{1}{x}\right| \text{ if } x < 0 \\ 1 \text{ if } x \geqslant 0 \end{Bmatrix} \quad (x \in R)$$

Exercise 5

(i) The graph of the equation

$$x^2 + y^2 = 16 \quad (x, y \in R)$$

is a circle. Draw this circle and indicate the region representing the solution set of the inequality

$$x^2 + y^2 < 16 \quad (x, y \in R)$$

(ii) Illustrate the solution set of

$$2x - 3y < 4 \quad (x, y \in R)$$

and also of

$$2x - 3y \leqslant 4 \quad (x, y \in R)$$

1.8 Answers to Exercises

Section 1.2

Exercise 1

(i) (a), (e), (f), (g) are correct.
(ii) None of the statements is correct.

The fact that A and B have the same number of elements is not enough to give equality. For two sets to be equal, not only must they have the same number of elements but these elements must be the same.

Section 1.3

Exercise 1

(i) 75	(v) 240
(ii) 51	(vi) 10
(iii) 1944	(vii) $\{75, 120, 175\}$
(iv) 20	

Exercise 2

(i) FALSE
A function is a special type of mapping. For example if A is the set of all television receivers in Great Britain and B is the set of all their valves, the rule which assigns to each television receiver its valves is a mapping, but not a function, because each receiver maps to numerous valves. But if we took as codomain the set of all television trade names, then the mapping of receivers to trade names is a function because each receiver has a unique trade name.

(ii) FALSE

(iii) TRUE
By definition, the codomain contains all the images of the elements of the domain.

(iv) FALSE
The codomain may contain elements which are not images under the mapping.

(v) TRUE
The list satisfies the requirements of a "mapping rule". It assigns at least one element of B to *each* element of A.

(vi) FALSE
The mapping is not a function because α does not have a *unique* element as its image.

(vii) FALSE
No element is assigned to γ.

Exercise 3

(i) (a) R^+ (b) $\{3, 5, 7\}$ (c) The set of real numbers greater than 1

(ii) (a) R (b) $\{-1, 2, 7\}$ (c) The set of real numbers greater than or equal to -2

(iii) (a) R (b) 3 (c) 3

This last function is an example of a constant function. A constant function is one for which the image of every element of the domain is the same.

Section 1.4

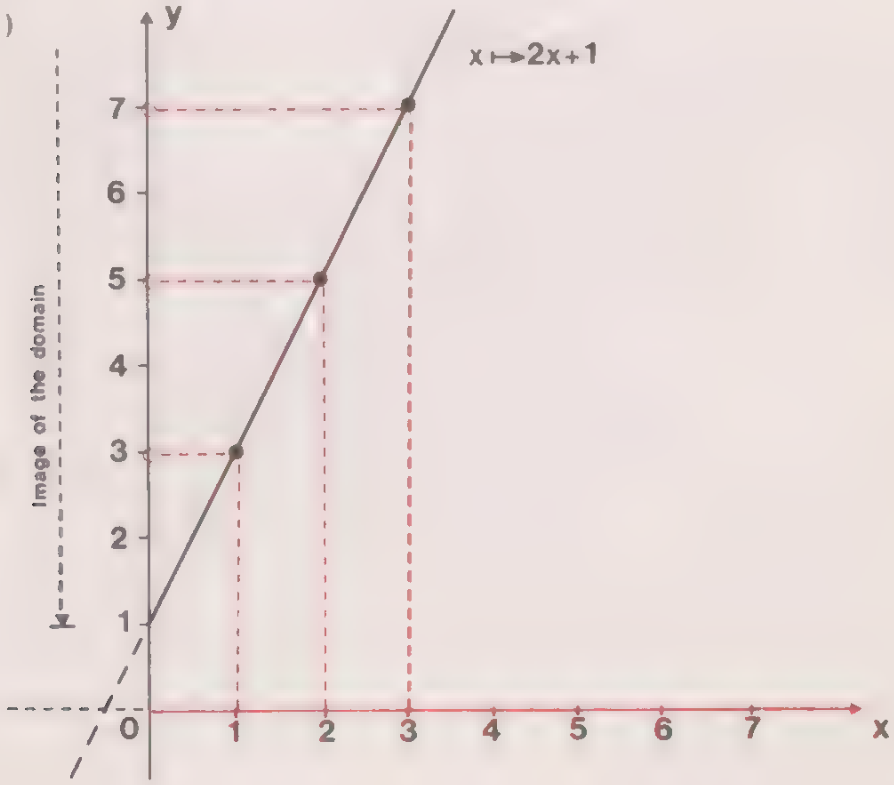
Exercise 1

A function must specify only one image for each element of the domain, and so the graph of a function must do likewise — any line drawn parallel to the "codomain axis" must cut the curve at most once. Hence (c) does not specify a function.

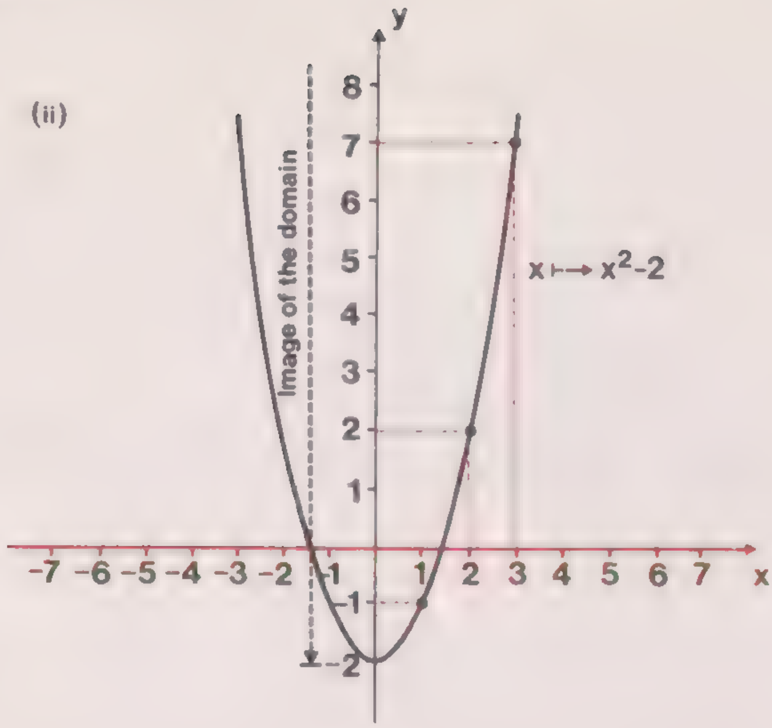
(a) Function
(b) Function
(c) Not a function
(d) Function

Exercise 2

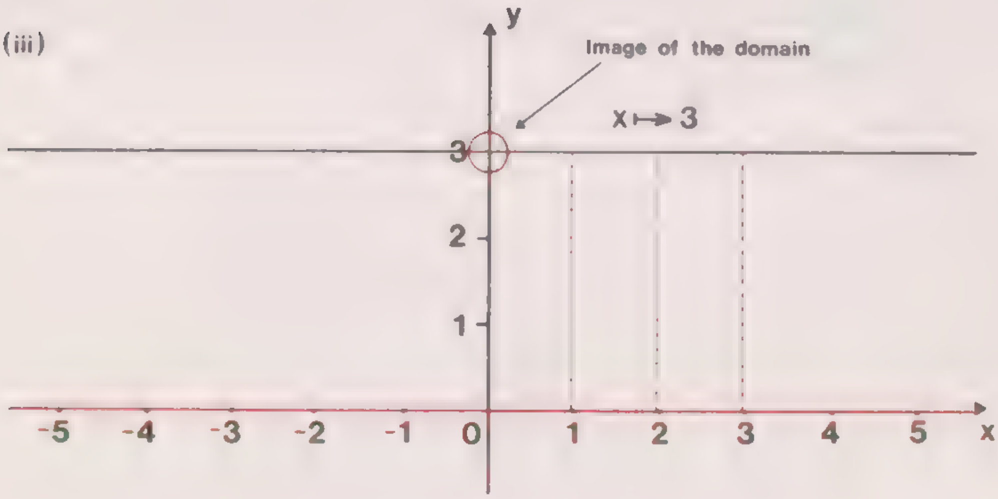
(i)



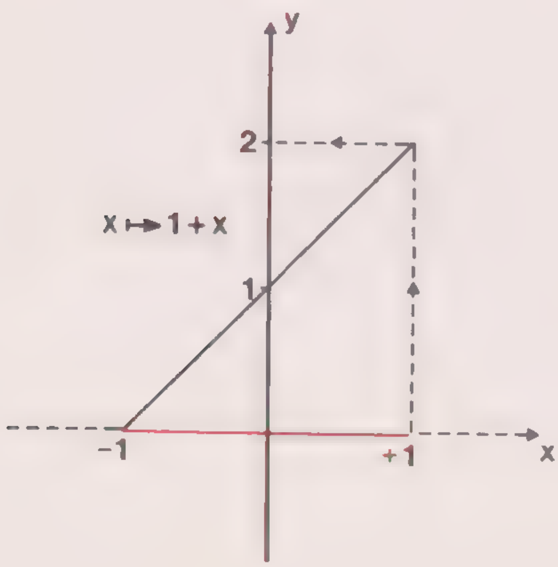
(ii)

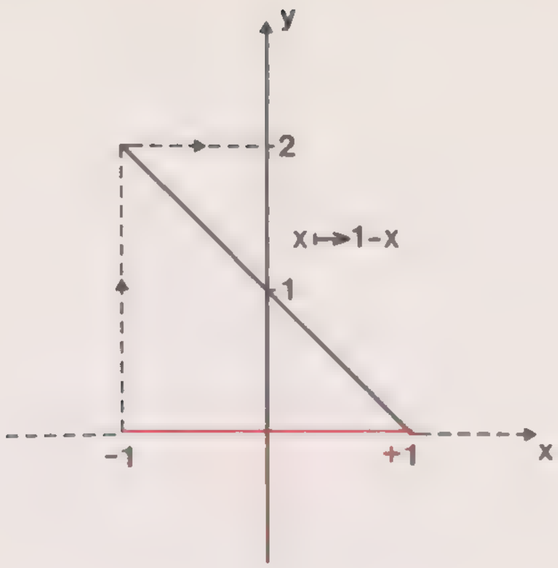


(iii)



Exercise 3





Section 1.5

Exercise 1

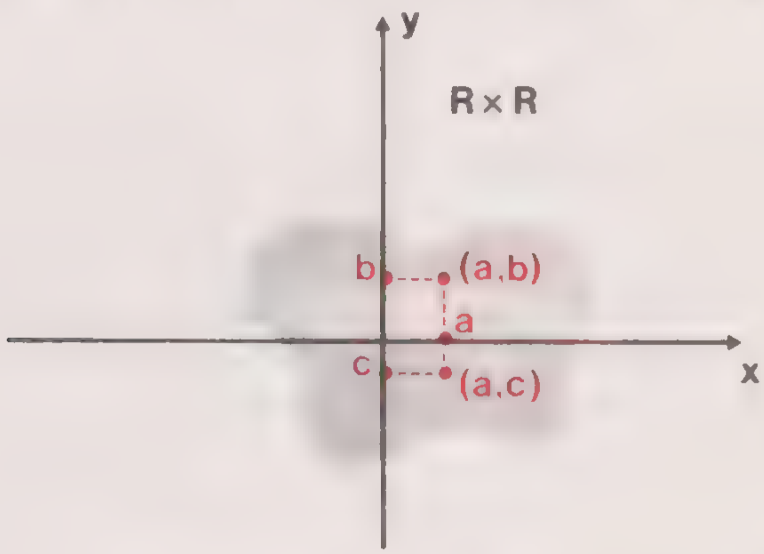
(i), (iii), and (iv) define mappings; (ii) does not, because no images are assigned to the elements b and c.

The mappings (i) and (iv) define functions.

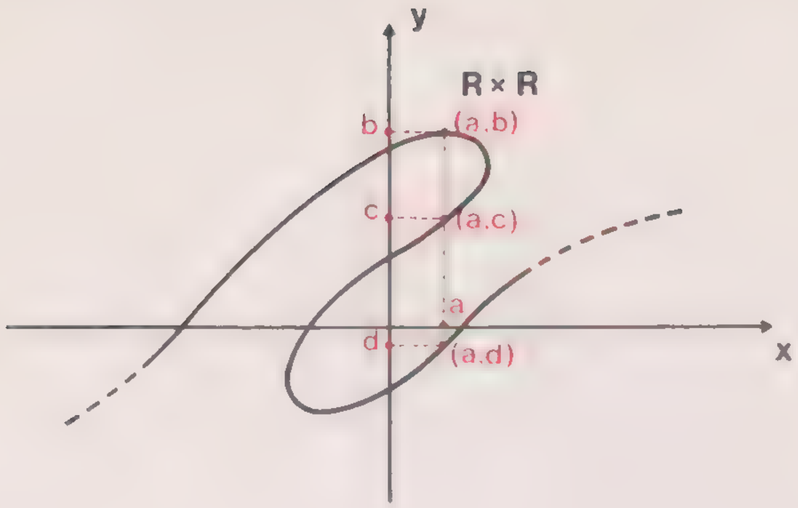
The mapping (iii) does not define a function because, for instance, the element a does not have a unique element as its image.

Exercise 2

No *area* can define a function $R \longrightarrow R$. If the ordered pair (a, b) belongs to the area then there are other ordered pairs (a, c) with $c \neq b$ also belonging to the area, and a thus has more than one element in its image.



Similarly, in the case of the second subset illustrated, there will be elements such as a having more than one element in its image because of the way in which the curve "folds back" on itself.



Section 1.6

Exercise 1

(i)

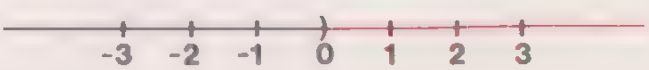
(ii)

(iii) $\{x : x \in R, x^2 + 3x + 2 = 0\} = \{-1, -2\}$

(iv)

(v)

In cases (iv) and (v) it is not clear from the illustrations whether or not the respective end-points, 0 and 2, are included in the solution sets. There is a way of indicating this which is often used. To indicate that 0 is excluded in case (iv) we draw the following diagram:



To indicate that 2 is included in case (v) we draw:

Similarly, the diagram:

represents the set $\{x : x \in R, 1 \leqslant x < 6\}$ in which 1 is included and 6 is excluded.

Exercise 2

(i) $N = 3$ will serve the purpose. One can find this number, and many others, by trial and error.

(ii) As n gets larger and larger, $\dfrac{1}{n+1}$ gets smaller and smaller. If we can find *any* one value of N such that $\dfrac{1}{N+1} < 0.01$, then we know that $\dfrac{1}{n+1} < 0.01$ for all values of n greater than N. One such value for N is

$$N = 100$$

(iii) $\delta = 0.3$ is a possibility.

Section 1.7

Exercise 1

(i) FALSE
(ii) TRUE
(iii) FALSE
(iv) TRUE
(v) FALSE
(vi) FALSE
(vii) TRUE
(viii) FALSE

Exercise 2

(i) TRUE
(ii) FALSE
(iii) TRUE ($m(a)$ stands for the image of a under m. The image belongs to the set B, and so $m(a) \in B$.)

Exercise 3

(i) TRUE
The statement defines f completely; 2 is in the domain and so we can substitute in the formula, and $2 \times 2 + 1 = 5$.

(ii) FALSE

The statement does not define f because the domain is not specified and so we do not know whether we are allowed to substitute the number 1 into the formula.

(iii) FALSE

f could be one of any number of functions which include 2 in the domain and which map 2 to 5, e.g.

$$f: x \longmapsto x^2 + 2x - 3 \quad (x \in R)$$

(iv) FALSE

-10 is not in the domain of f.

(v) TRUE

Since the domain is R, we *can* substitute any real number into the formula.

(vi) TRUE

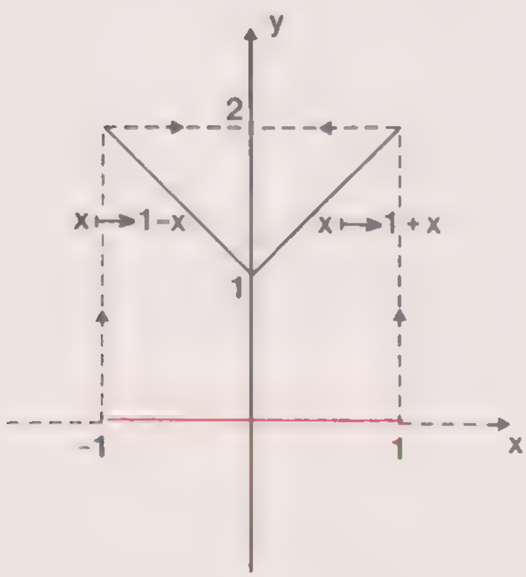
This is an example to illustrate the fact that any letter can be used as a variable in the definition of a function. Sometimes such a letter is called a *dummy variable*.

(vii) FALSE

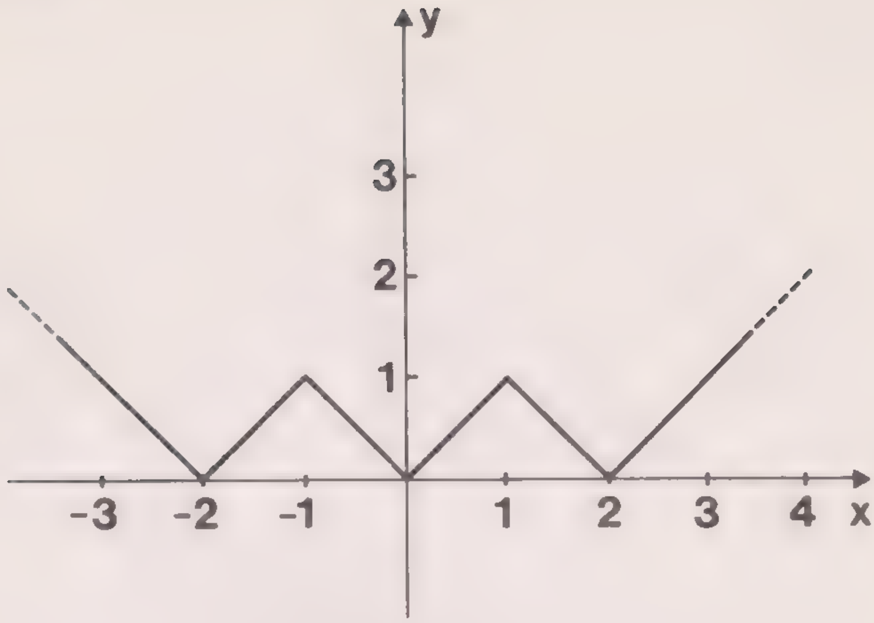
The domains are not the same.

Exercise 4

(i)



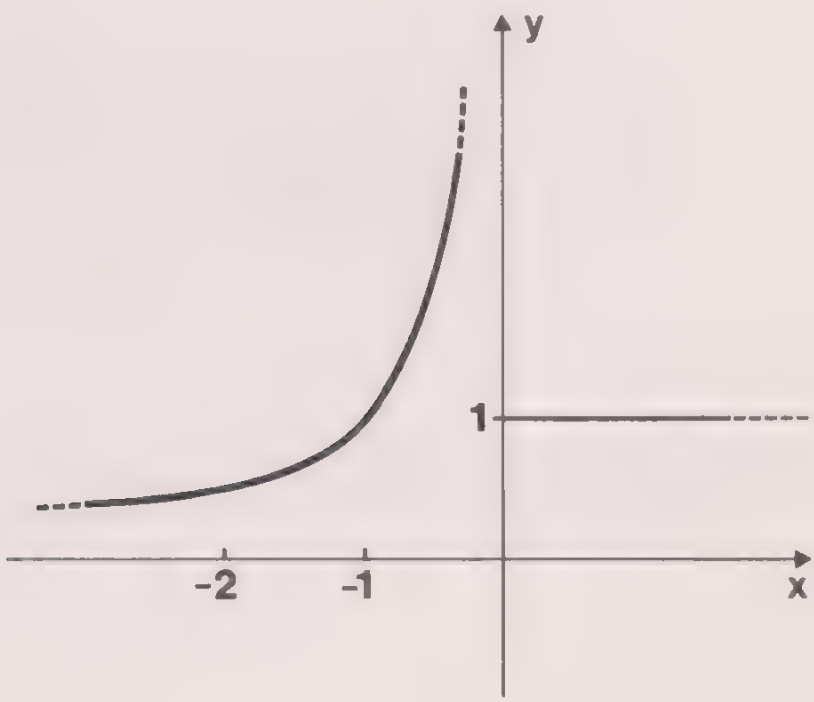
(ii)



This is a rather tricky one. The straight lines are

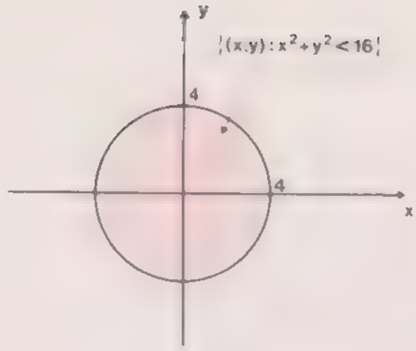
$$
\begin{array}{ll}
x \longmapsto -x-2 & (x \leqslant -2) \\
x \longmapsto x+2 & (-2 \leqslant x \leqslant -1) \\
x \longmapsto -x & (-1 \leqslant x \leqslant 0) \\
x \longmapsto x & (0 \leqslant x \leqslant 1) \\
x \longmapsto 2-x & (1 \leqslant x \leqslant 2) \\
x \longmapsto x-2 & (x \geqslant 2)
\end{array}
$$

(iii)

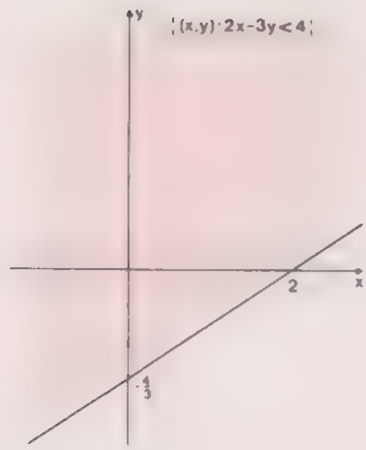


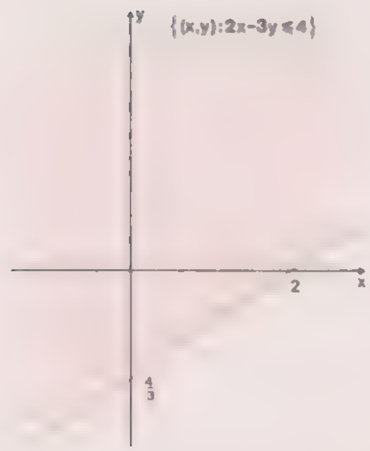
Exercise 5

(i)



(ii)





This time the set of points represented by the line $2x - 3y = 4$ *is* included in the solution set, so we have shown the line in red.

CHAPTER 2 SEQUENCES

2.0 Introduction

In this chapter we are concerned with the concept of *sequence*, and we use ideas from Chapter 1 to present it.

At first, we concern ourselves with finite sequences, but, once the concept of sequence has been sufficiently illustrated, we go on to consider infinite sequences.

Infinite sequences lead on naturally to the concept of *limit*, and this idea is introduced from an intuitive standpoint only, a more rigorous approach being deferred until later.

2.1 What is a Sequence?

The idea of a sequence is one that is frequently encountered in everyday experience, for example, a sequence of events such as a sequence of operations performed in the manufacture of some article.

A sequence is formally defined as follows

> A SEQUENCE is a collection of objects (not necessarily all different) arranged in a definite order.

Some examples of sequences are:

(i) A line of four cars waiting at a red traffic light.

(ii) The six words forming the sentence:

"Division by zero is not defined"

taken in the order in which they occur:

division, by, zero, is, not, defined.

(iii) The ten numbers from 1 to 10 taken in their natural order:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

The objects forming the sequence (cars, words or numbers in these examples) are called its elements. For the present we shall consider only finite sequences like those above; that is, sequences comprising a definite number of elements. Infinite sequences will be considered later.

There are several ways of specifying a sequence; the simplest way, used in our examples above, is to list all the elements in order. Sometimes we use an *incomplete list* and indicate by dots that there are missing elements; for example, we could abbreviate the third list to:

$$1, 2, 3, \ldots, 10$$

Such incomplete lists should be used only in cases where the context makes it clear what the missing elements are.

The notation used for listing sequences is similar to the notation for sets, but we distinguish sequences from sets by enclosing the lists in braces in the case of a set only. This notational distinction is necessary because of the distinction between the concepts of a sequence and a set, which is that the order of the elements matters in a sequence but not in a set. Re-arranging the elements in a sequence gives a new sequence, but re-arranging the elements in a set gives the same set. For example, the sequence 1, 2, 3 is distinct from the sequence 3, 1, 2 but

$$\{1, 2, 3\} = \{3, 1, 2\}$$

However, it may not always be practicable to describe a sequence by means of a complete or incomplete list. For example, if a sequence has a million elements, then a complete list of them may occupy a thousand pages, and an incomplete list may not give enough information to specify the sequence unambiguously. In such cases it may be possible to describe the sequence economically by giving a rule or formula for determining which object appears in each position of the sequence. (In fact, some sequences occur naturally this way, as we shall see.) In the language of Chapter 1, we specify the sequence by defining a function. The domain of this function will comprise the first N natural numbers, i.e. be the set $\{1, 2, 3, \ldots, N\}$, where N is the number of elements in the sequence; the rule must be one from which, given any natural number k belonging to the domain, we can determine the kth member of the sequence. As a simple example, the sequence comprising the reciprocals of the first million natural numbers is specified by the function, f say, with rule $k \longmapsto \frac{1}{k}$ and domain $\{1, 2, \ldots, 1\,000\,000\}$.

We may specify f by means of the formula

$$f: k \longmapsto \frac{1}{k} \qquad (k \in \{1, 2, \ldots, 1\,000\,000\})$$

or

$$f(k) = \frac{1}{k} \qquad (k \in \{1, 2, \ldots, 1\,000\,000\})$$

Either of these formulas tells us that, for every integer k between 1 and 1 000 000, the kth term of the sequence, denoted by $f(k)$, is equal to $\frac{1}{k}$.

As a further abbreviation, it is customary to write u_k (or some other letter with subscript k) rather than $f(k)$ for the kth element of the sequence. It is also customary to abbreviate the description of the domain slightly, so that the above example would usually be shortened to

$$u_k = \frac{1}{k} \qquad (k = 1, 2, \ldots, 1\,000\,000)$$

Finally, we can use a *recurrence formula* which expresses each element of a sequence in terms of one or more of its predecessors. But now we must also state the first element(s) and thus give the sequence a fair start.

Examples are

$$u_k = \frac{u_{k-1}}{k} \qquad (u_1 = 1;\, k = 2, \ldots, 1\,000\,000)$$

$$u_k = \frac{1}{2}\left(u_{k-1} + \frac{a}{u_{k-1}}\right) \quad (u_1 = 1;\, k = 2, \ldots, 20)$$

The most general recurrence formula would have the form

$$u_k = F(u_{k-1}, u_{k-2}, u_{k-3}, \ldots)$$

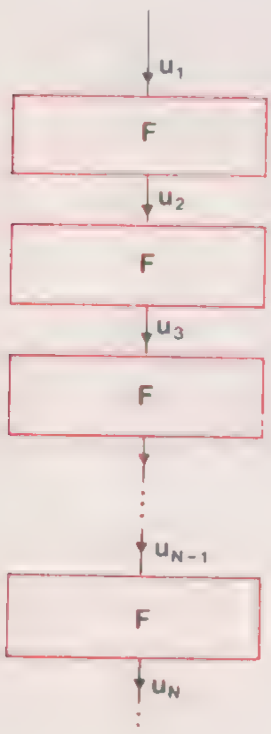
where F is some function. At this time we shall only consider the simplest type of recurrence formula, in which each element depends only on its immediate predecessor. A general recurrence formula of this type is

$$u_k = F(u_{k-1})$$

where F is some function. The calculation implied by this formula can be represented by a diagram:

indicating that we put the number u_{k-1} into the function F and get out of it the number u_k.

Since k can have any of the values $2, 3, \ldots, N$, where N is the number of elements in the sequence, the above diagram really stands for $(N - 1)$ different diagrams in which k takes these different values. These $(N - 1)$ diagrams can be joined up to give a new diagram representing the entire process by which we compute the successive elements of the sequence $u_1, u_2, \ldots, u_N$.



Exercise 1

Write the sequence specified by

$$u_n = n(n+1) \quad (n = 1, 2, \ldots, 5)$$

as a complete list.

Specify it in another way, using a recurrence formula.

2.2 Infinite Sequences

In the previous section we restricted our discussion to finite sequences, i.e. sequences with a finite number of elements. Thus a finite sequence can be specified by a function $k \longmapsto u_k$ with domain $\{1, 2, \ldots, N\}$. On the other hand a sequence specified by a function $k \longmapsto u_k$ with domain Z^+, the set of *all* positive integers, is called an infinite sequence.

How we Specify Infinite Sequences*

To be able to discuss infinite sequences we must first be able to specify them. The methods used are just the same as for finite sequences except that a complete list is never possible. Here is an example where the same sequence is specified by the three different methods:

(i) incomplete list

$$1, 2, 4, 8, 16, 32, \ldots$$

(ii) function

$$k \longmapsto 2^{k-1} \quad (k \in Z^+)$$

(iii) recurrence formula

$$u_1 = 1$$

$$u_k = 2u_{k-1} \quad (k = 2, 3, \ldots)$$

Exercise 1

List the first 5 elements of the sequences specified by:

(i) $u_n = (-1)^n \qquad (n \in Z^+)$

(ii) $u_1 = 3$

$$u_k = 3 + \frac{u_{k-1}}{10} \quad (k = 2, 3, 4, \ldots)$$

Exercise 2

Write down functions that specify the following sequences:

(i) $1, -2, 3, -4, 5, -6, \ldots$

(ii) $u_1 = 0$

$$u_k = \frac{1}{2 - u_{k-1}} \quad (k = 2, 3, 4, \ldots)$$

Exercise 3

Write down specifications in terms of recurrence formulas for the sequences:

(i) $1, -1, 1, -1, \ldots$

(ii) $u_k = \dfrac{1}{(-2)^k} \quad (k \in Z^+)$

2.3 Limit of a Sequence

We are now in a position to take a preliminary look at the concept of a *limit*. As with any mathematical concept, there are two ways of looking at it: the intuitive and the rigorous. The intuitive aspect enables us to recognize the situations where the concept is likely to be useful, and the rigorous aspect enables us to apply it correctly. Both are essential to a proper understanding; it is true that many users of mathematics do succeed in getting by on the intuitive aspect alone, but it is rather like travelling in a car without a spare tyre: at any moment a situation may arise with which the available equipment cannot cope. We shall be discussing both aspects of the concept. Here we begin the intuitive approach, and we continue in Chapter 4. The rigorous definition will finally appear in Chapter 6.

If an infinite sequence is a sequence of successive approximations to some number, then we call that number the limit of the sequence. For example, the sequence

$$0.3, 0.33, 0.333, 0.3333, 0.33333, \ldots$$

is a sequence of successively closer decimal approximations to the number $\frac{1}{3}$: its limit is therefore $\frac{1}{3}$. If we denote the sequence u_1, u_2, u_3 by u, a convenient formulation of this intuitive notion is

Intuitive Definition of a Limit

"The number $\lim \underset{\sim}{u}$ is the limit of the infinite sequence $\underset{\sim}{u}$" is equivalent to the statement "if k is very large, then u_k is a very good approximation to $\lim \underset{\sim}{u}$".

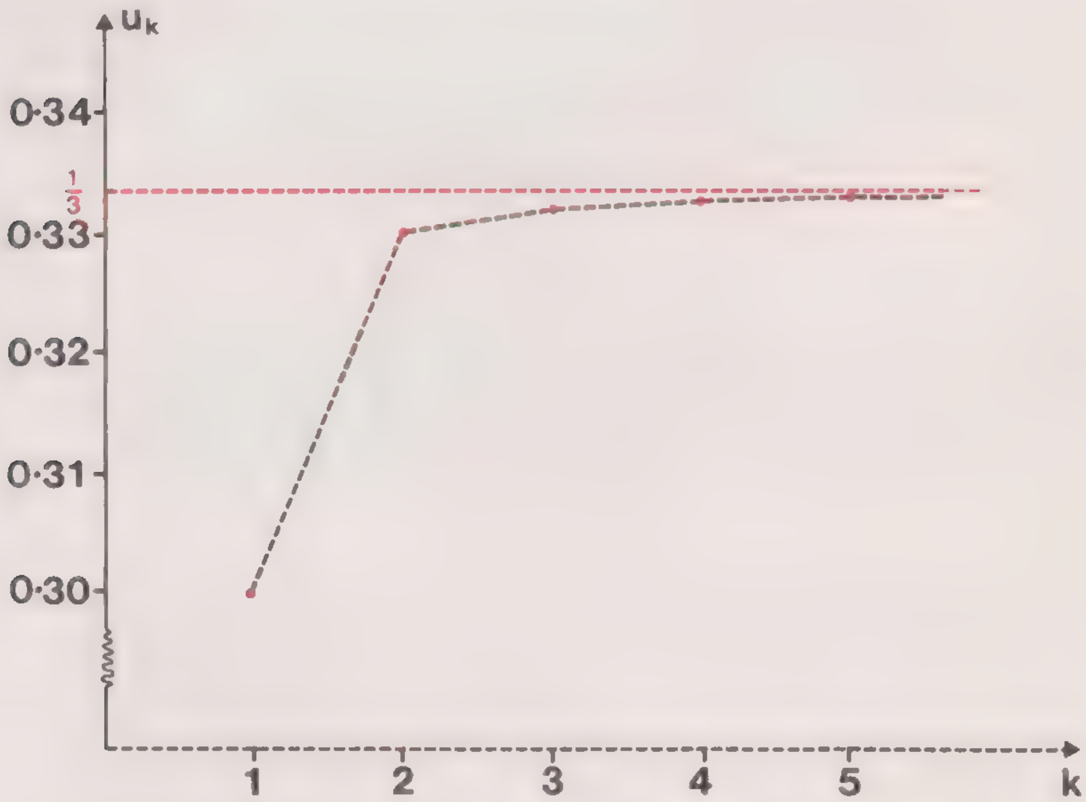
Not every sequence has a limit; for example neither of the sequences

$$1, 2, 4, 8, 16, \ldots, 2^{k-1}, \ldots \quad \text{and} \quad 1, 0, 1, 0, 1, \ldots$$

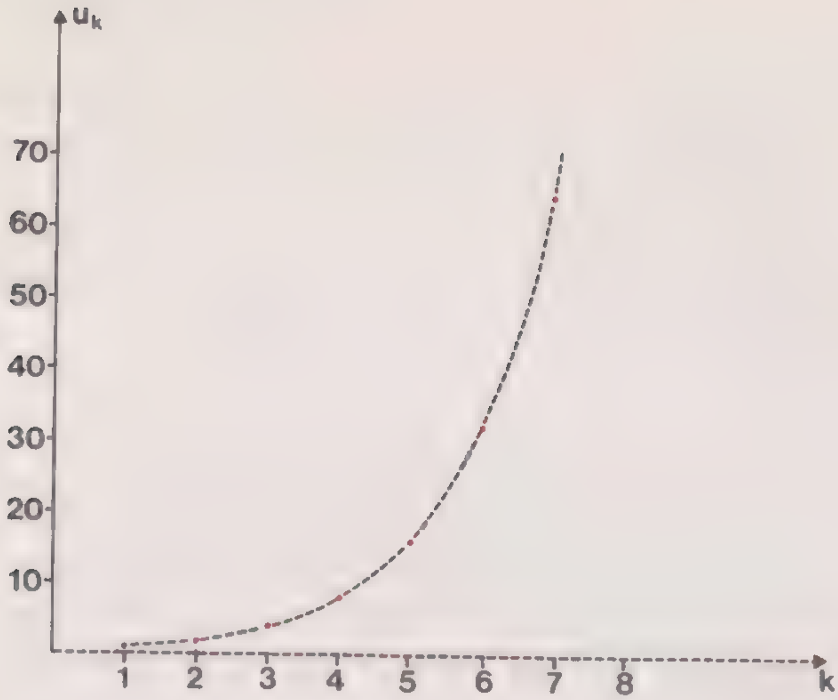
has a limit.

In the first sequence the elements increase with k and in the second they oscillate between 0 and 1: in neither case is there a number which satisfies our intuitive definition of a limit. We distinguish two types of sequence: we say that a sequence having a limit is convergent (or that it converges) and that one without a limit is non-convergent. (The term "divergent" is also very common.)

To see whether a sequence is convergent or not it is often helpful to look at its graph. Here are the graphs of the first two sequences given above. Can you see the geometrical property of the first graph that corresponds to the convergent character of the sequence?



Graph of the sequence 0·3, 0·33, 0·333, ...
The dashed lines are to guide the eye only; they are not part of the graph.

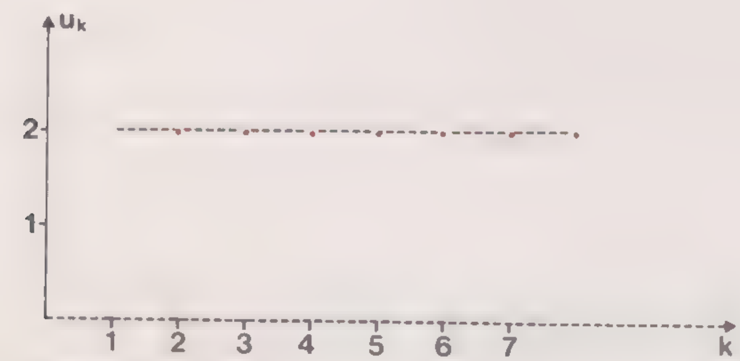


Graph of the sequence 1, 2, 4, 8, 16, 32, ...

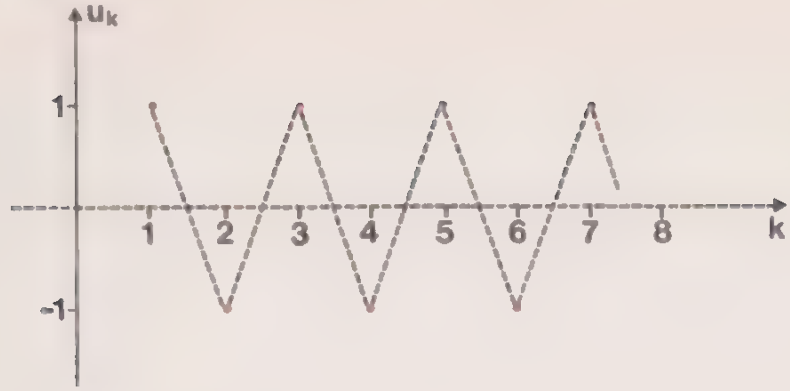
The first sequence consists of successive approximations to the number $\frac{1}{3}$, and so as k increases the points of the graph get steadily closer to the red line. In the second graph we cannot draw a line with this property. These examples illustrate that the graph of a convergent sequence $\underset{\sim}{u}$ is characterized by this property: for large k, the points (k, u_k) are very close to a line parallel to the k-axis and at a distance $\lim \underset{\sim}{u}$ from it.

Exercise 1

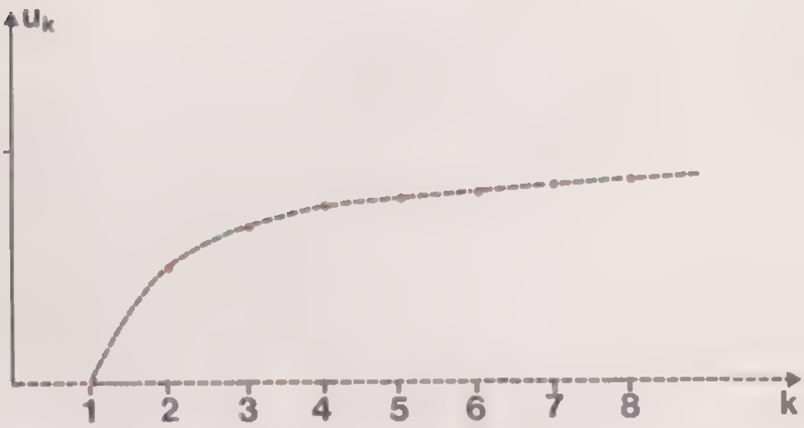
Here are graphs of some infinite sequences. Which are convergent and what are the limits of the convergent ones? ($k \in Z^+$ throughout.)



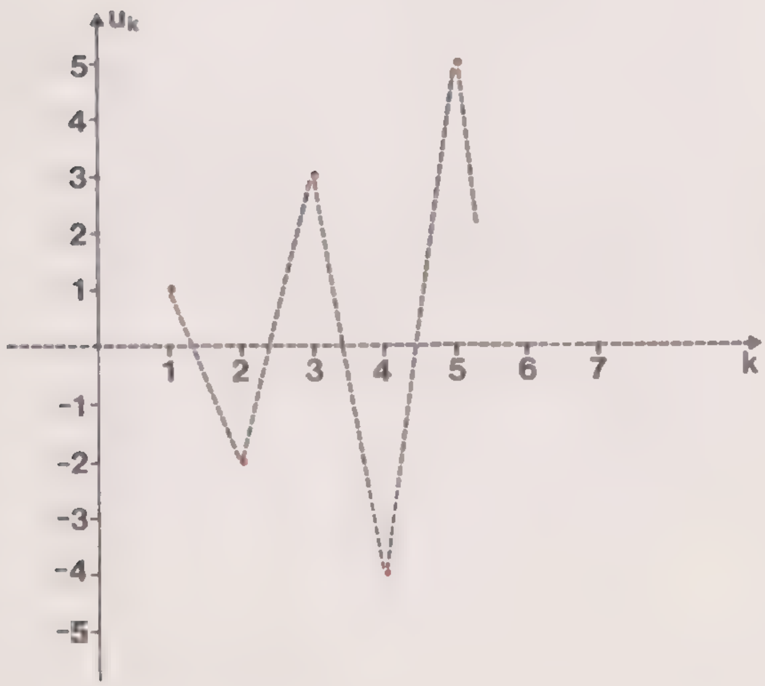
(i) Graph of $u_k = 2$



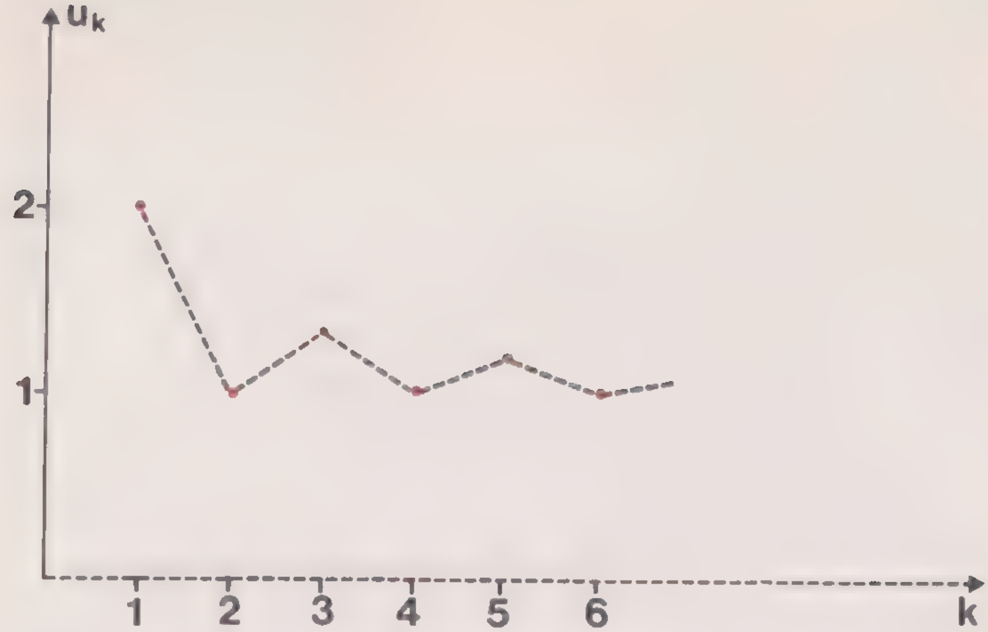
(ii) Graph of $u_k = (-1)^{k+1}$



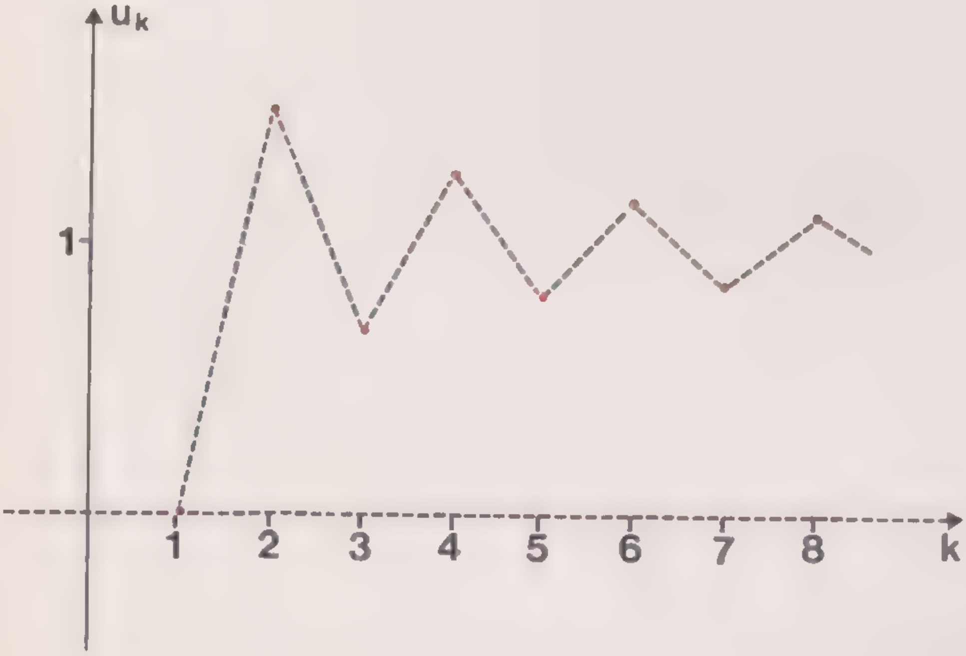
(iii) Graph of $u_k = \frac{k-1}{k}$



(iv) Graph of $u_k = (-1)^{k-1} k$



(v) Graph of $u_k = \begin{cases} 1+\frac{1}{k} & \text{if } k \text{ is odd} \\ 1 & \text{if } k \text{ is even} \end{cases}$



(vi) Graph of $u_k = 1 + \frac{(-1)^k}{k}$

Exercise 2

Which of the following sequences are convergent, and what are the limits of the convergent ones?

(i) $u_k = k$

(ii) $u_k = \dfrac{1}{k}$

(iii) $u_k = \begin{cases} \dfrac{1}{\sqrt{k}} & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}$

(iv) $u_k = 1 + \dfrac{1}{k}$

(v) $u_k = \dfrac{1}{10^6} + \dfrac{1}{k}$

where $k \in Z^+$ in each case.

This last exercise may give some idea of the difficulties we can get into if we "travel without a spare tyre" by relying entirely on the intuitive notion of a limit. For large k, the elements of the sequence $u_k = \dfrac{1}{10^6} + \dfrac{1}{k}$ in (v) are all very good approximations to the number $\dfrac{1}{10^6}$, and so $\dfrac{1}{10^6}$ satisfies the intuitive definition of the limit; but since $\dfrac{1}{10^6}$ is very close to 0, the elements are also "very good approximations" to 0, so that it would appear that 0 could equally well be called the limit. This shows that the intuitive definition of a limit given on page 12 can lead to ambiguities if it is pushed too far.

2.4 Additional Exercises

Exercise 1

Write down functions that specify the following sequences:

(i) $3, \frac{4}{3}, 1, \frac{6}{7}, \frac{7}{9}, \frac{8}{11}, \ldots$

(ii) $\frac{1}{2}, \frac{2}{5}, \frac{3}{10}, \frac{4}{17}, \frac{5}{26}, \ldots$

(iii) $\frac{1}{101}, \frac{2}{51}, \frac{9}{103}, \frac{2}{13}, \frac{5}{21}, \frac{18}{53}, \ldots$

Exercise 2

Determine in any way you can whether each of the sequences (i) — (iii) of Exercise 1 is or is not convergent. What are the limits of the convergent sequences?

2.5 Answers to Exercises

Section 2.1

Exercise 1

2, 6, 12, 20, 30

$$u_n = \frac{n+1}{n-1} u_{n-1} \quad (u_1 = 2; n = 2, \ldots, 5)$$

Section 2.2

Exercise 1

(i) $-1, 1, -1, 1, -1$

(ii) 3, 3.3, 3.33, 3.333, 3.3333

Exercise 2

(i) $k \longmapsto k \cdot (-1)^{k+1} \qquad (k \in Z^+)$

(ii) The sequence is

$$0, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{4}{5}, \tfrac{5}{6}, \ldots$$

and a function which specifies this sequence is

$$k \longmapsto \frac{k-1}{k} \quad \text{or } k \longmapsto 1 - \frac{1}{k} \quad (k \in Z^+)$$

Exercise 3

(i) $u_1 = 1$

$u_k = -u_{k-1} \qquad (k = 2, 3, 4, \ldots)$

(ii) $u_1 = -\frac{1}{2}$

$u_k = -\frac{1}{2} u_{k-1} \qquad (k = 2, 3, 4, \ldots)$

Section 2.3

Exercise 1

(i) Convergent, with limit 2. Here the points (k, u_k) all lie *on* the line parallel to the k-axis at a distance 2 units from it.
(ii) Not convergent.
(iii) Convergent, with limit 1.
(iv) Not convergent.

(v) Convergent, with limit 1.
(vi) Convergent, with limit 1.

Exercise 2

The quickest way to do these is by using the Intuitive Definition on page 49, but if you are unsure of how to use it, draw the graphs as well.

(i) Divergent. The elements increase with k; they never get close together, as they would if the sequence converged.
(ii) Convergent with limit 0. The elements get closer and closer to 0 as k increases.
(iii) Convergent with limit 0. Half the members of the sequence are actually equal to 0, and the other members get closer and closer to 0 as k increases.
(iv) Convergent with limit 1. Since $\frac{1}{k}$ is very small for large k, the quantity $1 + \frac{1}{k}$ is very close to 1 when k is large.
(v) Convergent with limit $\frac{1}{10^6}$. See the discussion on page 53.

Section 2.4

Exercise 1

(i) $k \longmapsto \dfrac{k+2}{2k-1} \qquad (k \in Z^+)$

(ii) $k \longmapsto \dfrac{k}{k^2+1} \qquad (k \in Z^+)$

(iii) $k \longmapsto \dfrac{k^2}{k+100} \qquad (k \in Z^+)$

Exercise 2

It is possible to obtain an indication of convergence (or otherwise) by drawing graphs, but a little very simple algebraic manipulation is better because the sequences do not necessarily converge to their limits very quickly.

(i) Rewrite as

$$k \longmapsto \frac{1+\dfrac{2}{k}}{2-\dfrac{1}{k}} \qquad (k \in Z^+)$$

by dividing numerator and denominator by k. As k becomes very large, $2/k$ and $1/k$ become very small, and the sequence will converge with limit 1/2.

(ii) Rewrite as

$$k \longmapsto \frac{1}{k + \frac{1}{k}} \qquad (k \in Z^+)$$

As k becomes very large, $1/k$ becomes very small and the sequence will converge with limit 0.

(iii) Rewrite as

$$k \longmapsto \frac{k}{1 + \frac{100}{k}} \qquad (k \in Z^+)$$

As k becomes very large, $100/k$ becomes very small and we are left with k. The sequence therefore does not converge.

CHAPTER 3 FUNCTIONS

3.0 Introduction

In this chapter we continue the discussion started in Chapter 1 and consider those mappings which are also *functions*.

We begin with a classification of mappings and functions and then we consider the *arithmetic* of functions by analogy with the arithmetic of numbers.

We also consider the idea of a *function of a function* which involves a new way of combining functions known as *composition*.

It frequently happens that we want to reverse a mapping, that is, start with elements which are in the codomain. If the mapping is a function then we want to know under what circumstances the reverse of the mapping is a function. This problem is discussed in the section on *Inverse Functions*.

3.1 Kinds of Mapping

In Chapter 1 we gave the following definition of a mapping.

> A MAPPING consists of a set *A*, a set *B* and a rule by which an element (or set of elements) of *B* is assigned to *each* element of *A*.

Notice that an image must be assigned to *each* element of the domain.

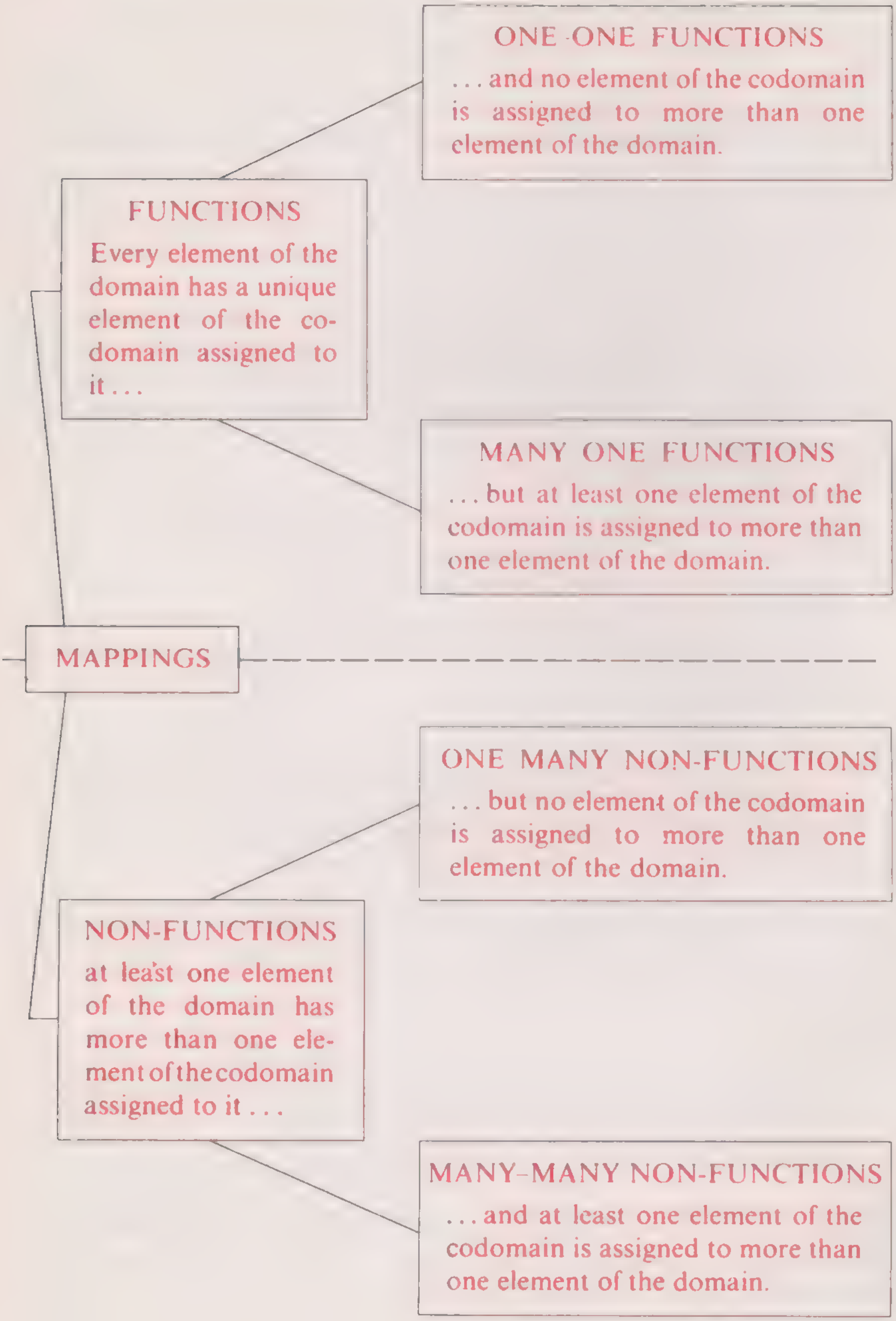
We also saw that the set of all the images is a subset of the codomain of a mapping.

It is possible to assign images in various ways. In particular, each element in the domain of a mapping may have only one element of the codomain as its image, in which case, the mapping is a function.

> A FUNCTION is a mapping for which each element in the domain has only one element as its image.

Again, there are some functions where no element of the codomain is assigned to more than one element of the domain, and others where several elements of the domain have the same image.

The various possibilities are summarized in the following diagram:



We shall now be concerned only with those mappings which are functions.

Exercise 1

Classify the following functions as

$\boxed{\text{one–one}}$ or $\boxed{\text{many–one}}$

(i) $x \longmapsto 3x^2 + 2 \qquad (x \in R)$
(ii) $x \longmapsto x^3 + 2 \qquad (x \in R)$
(iii) $x \longmapsto \sin x \qquad (x \in R)$

3.2 The "Arithmetic" of Functions

When one learns a new game, there are usually two stages to go through. First of all one has to find out something about the new materials — "take an ordinary pack of playing cards" or "take 22 men and a leather ball" or "take one egg, a cupful of flour and half a cupful of milk". The next step is to learn the rules of the game, that is to say, how to use the materials.

This is the stage we are at now. We have defined our materials, and we are about to learn the rules of the game.

Suppose two functions f and g have R as domain and codomain. Then we can define

$\boxed{\text{the SUM of } f \text{ and } g}$

which we write as $f + g$, by

$$\boxed{f + g : x \longmapsto f(x) + g(x) \quad (x \in R)}$$

Example 1

Let

$$f : x \longmapsto x^2 \qquad (x \in R)$$

and

$$g : x \longmapsto x^6 \qquad (x \in R)$$

Then

$$f + g : x \longmapsto x^2 + x^6 \quad (x \in R)$$

It is quite natural to define the other "arithmetical" operations as follows:

DIFFERENCE

$$f - g : x \longmapsto f(x) - g(x) \quad (x \in R)$$

PRODUCT

$$f \times g : x \longmapsto f(x) \times g(x) \quad (x \in R)$$

QUOTIENT

$$f \div g : x \longmapsto \frac{f(x)}{g(x)}$$

The specification of the domain of the quotient is not straightforward. This is because of the difficulty which occurs when $g(x) = 0$. In this case the image of x is undefined, and we must remove such elements from the domain. So the domain of $f \div g$ is R with these elements omitted.

Exercise 1

In the text above we started with "Suppose two functions f and g have R as domain and codomain". This is unnecessarily restrictive. Do both domain and codomain have to be R in order to define the operations of arithmetic used here? What must be true of the domains of f and g? Can you think of an example where either domain or codomain are not R?

Exercise 2

If the functions f and g are defined by

$$f : x \longmapsto 6x^2 \quad (x \in [-1, 1])$$

and

$$g : x \longmapsto 6x \quad (x \in [-1, 1])$$

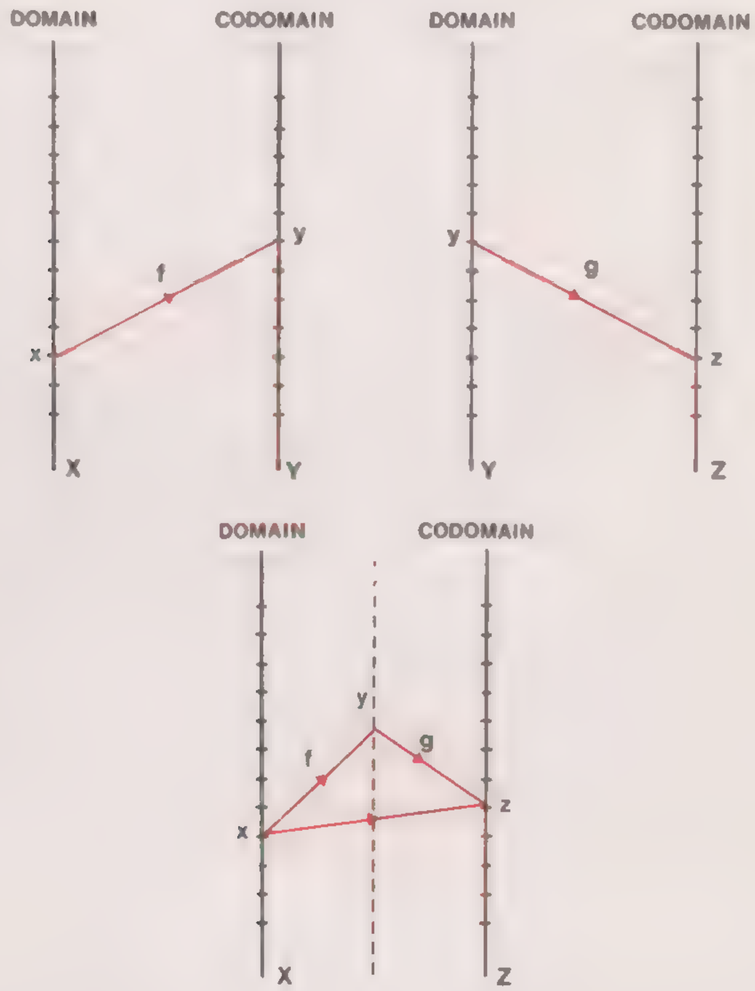
fill in the formula and the appropriate domain for

(i) $g + f$
(ii) $g \div f$
(iii) $f \div g$
(iv) $f \times g$

3.3 Composition of Functions

There is another way of combining functions which is fundamentally different from the "arithmetical" combinations of the last section. The emphasis of this combination is on the mapping from one set to another, rather than being a simple generalization of the operations of ordinary arithmetic.

Example 1



Notice that in the above example

$$f:\mathbf{x}\longrightarrow\mathbf{y} \quad \text{and} \quad g:\mathbf{y}\longrightarrow\mathbf{z}$$

The mapping in the bottom figure is obtained by using first f and then g. If we call it h, then

$$h:\mathbf{x}\longrightarrow\mathbf{z}$$

Example 2

Suppose that we have the functions

$$f = \boxed{\text{DOUBLE IT}} \text{ with domain } R$$

$$g = \boxed{\text{SQUARE IT}} \text{ with domain } R$$

then

$$x \mapsto \boxed{\begin{matrix} f \\ \text{DOUBLE IT} \end{matrix}} \to 2x \quad (x \in R)$$

and

$$x \mapsto \boxed{\begin{matrix} g \\ \text{SQUARE IT} \end{matrix}} \to x^2 \quad (x \in R)$$

Suppose now that we construct the function

$$x \mapsto \boxed{\begin{matrix} f \\ \text{DOUBLE IT} \end{matrix}} - 2x - \boxed{\begin{matrix} g \\ \text{SQUARE IT} \end{matrix}} \to 4x^2$$

by using *first* f and *then* g. If we call this *composite function* h, then

$$h : x \longmapsto 4x^2 \quad (x \in R)$$

Composition

These examples each illustrate the same method of composition of functions. Unlike the operations of the last section, which were simply extensions of the operations of ordinary arithmetic, there is no analogue of this new composition in number arithmetic. We cannot extend the use of a symbol as we did for $+$, $-$, $\times$ and $\div$ because no such symbol exists; so we must invent one. The symbol which is commonly adopted is the small circle $\circ$.

Thus $g \circ f$ (pronounced "gee oh eff") stands for the function obtained

by performing f *first*, and *then* g.

If we can define a function h by the rule:

$$h(x) = g(f(x))\ (x \in \text{domain of } f)$$

then we denote this function by

$$h = g \circ f$$

(Sometimes we refer to $g \circ f$ as a *function of a function*.)

Diagrammatically we have the extended rule

$$x \longmapsto f(x) \longmapsto g(f(x))$$

It is very important to notice that $g \circ f$ means that we use f first and then g

Example 3

If functions f and g are defined by

$$f : x \longmapsto 2x + 3 \quad (x \in R)$$

and

$$g : x \longmapsto x^2 - 1 \quad (x \in R)$$

we can calculate $g(f(x))$ by replacing x by $f(x)$ in the expression for $g(x)$:

$$g(x) = x^2 - 1$$

and so

$$g(f(x)) = [f(x)]^2 - 1$$

But

$$f(x) = 2x + 3$$

and so

$$\begin{aligned} g(f(x)) &= (2x + 3)^2 - 1 \\ &= 4x^2 + 12x + 8 \end{aligned}$$

Thus $g \circ f$ is the function defined by

$$g \circ f : x \longmapsto 4x^2 + 12x + 8 \quad (x \in R)$$

Exercise 1

(i) If f and g are functions defined by

$$f: x \longmapsto x - 1 \quad (x \in R)$$

and

$$g: x \longmapsto x^2 \quad (x \in R)$$

complete the following:

(a) $f \circ g : x \longmapsto ?\quad (x \in R)$
(b) $g \circ f : x \longmapsto ?\quad (x \in R)$

(ii) If f is the mapping which translates English to French and g is the mapping which translates French to German is it $g \circ f$ or $f \circ g$ which translates English to German?

Exercise 2

Given any two functions f and g,

(i) Can we always form $g \circ f$?
(ii) If we can form $g \circ f$, can we necessarily form $f \circ g$?
Explain your answers.

3.4 Decomposition of Functions

In the last section we were discussing methods of constructing formulas to describe composite functions. But it is often just as important to take a formula to pieces as to construct one. This is in fact just the process which is required when a formula is to be prepared for "digestion" by a computer.

Example 1

Consider the simple function

$$f: x \longmapsto 2x + 1 \quad (x \in R)$$

and suppose that we are asked for $f(7)$. We reply 15 with hardly any thought at all. But how would we describe the calculation to a machine?

INSTRUCTIONS

(i) Multiply the number which I shall give you by 2.
(ii) Add 1 to the result of (i).
(iii) Print out the result of (ii).

We could, for example, describe this function f as a composition of two simpler functions g and h

$$h : x \longmapsto 2x \qquad (x \in R)$$

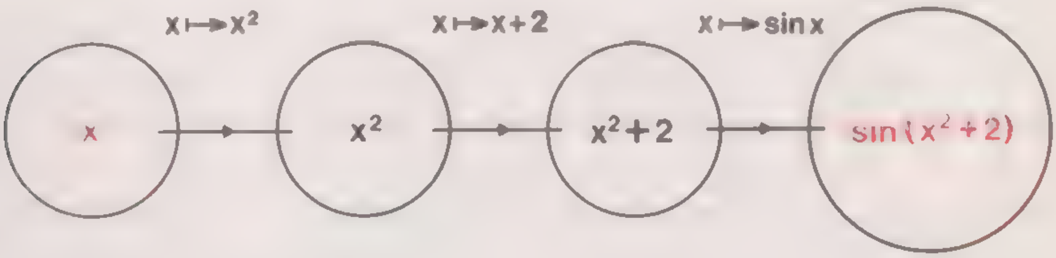
and

$$g : x \longmapsto x + 1 \qquad (x \in R)$$

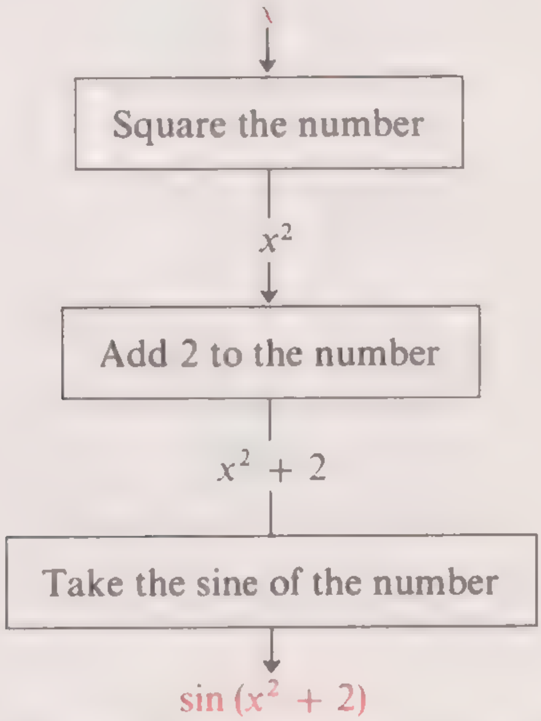
then

$$g \circ h : x \longmapsto 2x + 1 \quad (x \in R)$$

One can imagine that for complicated functions this break-down of the calculation into simpler steps is an important part of numerical work. It is often worthwhile in complicated calculations to plan the steps either by using a suitable mapping diagram, for example



or a FLOW DIAGRAM



The ability to break down a complicated procedure into small units, whether it be a calculation or a logical organizational problem, is important in many fields.

Exercise 1

Represent the break-down of the function

$$f:x \longmapsto \frac{\cos(x-\theta)}{2} \quad (x \in R)$$

by (i) a suitable mapping diagram

and

(ii) a flow diagram.

3.5 Inverse Functions

Reverse Mappings

One of the examples which we used in the first chapter of this volume was the mapping of a set of people to their blood groups. This example illustrates a type of mapping, which often occurs, where it makes sense to reverse the mapping. In this case, the mapping is reversed when somebody needs a blood transfusion, and a donor has to be found. In this section we shall discuss this idea of reversing a mapping.

If one is trying to locate a book in a library, then the recommended procedure is to look it up in the catalogue and find its classification number: the classification system provides a mapping, c say, such that

$$c:\text{Books} \longrightarrow \text{Classification Numbers}$$

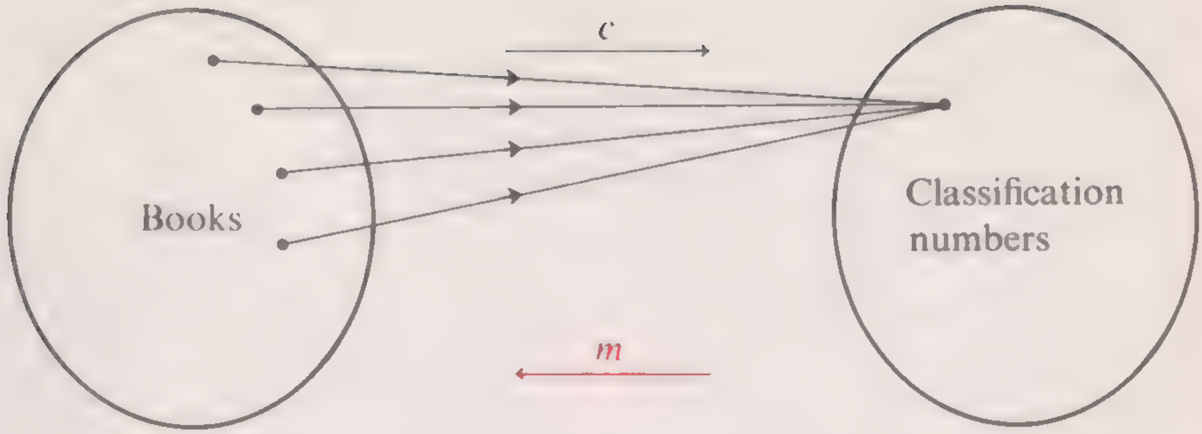
It is highly desirable that c is a function, rather than just a mapping. Can you say why?

On the other hand, if you want to find a book on a particular subject, and you know the classification number of the subject, then you look in a different catalogue which represents the mapping, m say, such that

$$m:\text{Classification Numbers} \longrightarrow \text{Books}$$

If m is a function rather than a mapping, then you can be sure that the library is not much good. Can you say why?

We say that m is the REVERSE MAPPING to c



A Definition

As we keep pointing out in this volume, it is not sufficient in mathematics just to understand a concept, we must be able to define it precisely. Probably the most convenient way to define a reverse mapping is in terms of the list of pairs which a mapping defines. A mapping f from A to B has a graph which is defined as the set of all pairs (x, y) such that $x \in A$ and y is $f(x)$ or, if $f(x)$ is a set of elements, belongs to $f(x)$.

> If f maps A to B and
>
> $$S = \{(x, y) : x \in A \quad \text{and} \quad y = f(x) \text{ or } y \in f(x)\}$$
>
> then the mapping of a subset of B to A, whose graph is
>
> $$\{(y, x) : (x, y) \in S\}$$
>
> is called the **reverse mapping** of f.

This definition is just a precise way of saying that we reverse the order of all pairs in the list which defines the mapping in order to get the reverse mapping. But one point is not entirely clear. What is the domain? We have said it is a subset of B. Consider the following examples.

Example 1

Consider the mapping

Set of Persons to Set of Telephone Numbers

A telephone directory is a list of all pairs

(Person, Telephone Number)

If we reverse the order, to get a list of all pairs

(Telephone Number, Person)

This would represent the reverse mapping.

Example 2

Let us look once again at our "Colour of eyes" example

Person living in England $\longmapsto$ Colour of his eyes

To reverse this we simply write

Colour of eyes $\longmapsto$ Person living in England with that colour of eyes

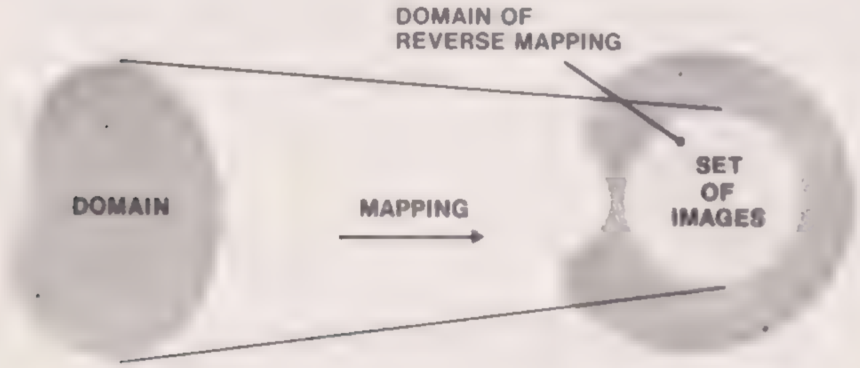
But we included CRIMSON in our list of colours and

CRIMSON $\longmapsto$?

Is the reverse mapping in this case truly a mapping? What is the domain of the reverse mapping? (Look carefully at the definition of MAPPING if you are not sure how to answer these questions.)

You may well say that these difficulties were caused by our foolishness in including colours such as crimson in the codomain in the first place. But how are we to know what colours we ought to put in it before we examine the colour of everyone's eyes?

> If g is the reverse mapping of f, and f has domain A, then the domain of g is $f(A)$



We have said previously that functions are important, but when is the reverse mapping a function? Certainly there is no guarantee that the reverse of a function is a function.

Example 3

$$f: x \longmapsto x^2 \quad (x \in [-1, 1])$$

is a function. The reverse of f

$$g: x \longmapsto \{\sqrt{x}, -\sqrt{x}\} \quad (x \in [0, 1])$$

is not a function according to our definition, because each number in $[0, 1]$ (except 0) has *two* separate square roots.

In section 3.1 we summarized the various kinds of mapping by means of a diagram. We can now consider each of the possibilities in turn, giving us:

A one–one mapping is a function whose reverse mapping is also a function.

A many–one mapping is a function whose reverse mapping is not a function.

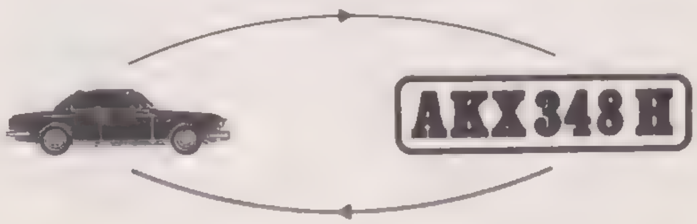
A one–many mapping is a mapping which is not a function, but whose reverse mapping is a function.

A many–many mapping is a mapping which is not a function, and whose reverse mapping is not a function.

Example 4

ONE–ONE MAPPING (FUNCTION)

The above mapping is ONE–ONE (at any rate the authorities try to make sure that it is.)

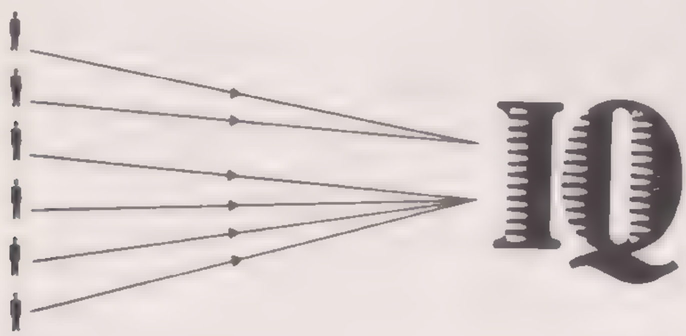


Each car has just one registration number. Each registration number corresponds to just one car.

Example 5

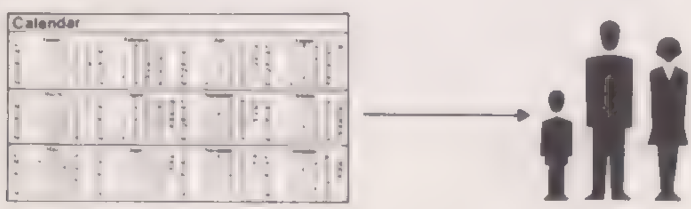
MANY–ONE MAPPING (FUNCTION)

The mapping from the set of all people in Great Britain to the set of integers obtained by mapping a person to his intelligence quotient (on Jan. 1st 1971 for example) is MANY ONE. Each person has a unique (i.e. just one) intelligence quotient (supposedly), but a large number of people map to 100, for example.



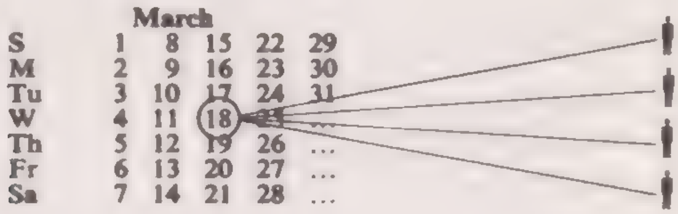
Example 6

ONE–MANY MAPPING (NOT A FUNCTION)



The mapping from the calendar to the set of all people living in Europe, obtained by mapping a date to the people born on that day, is ONE-MANY.

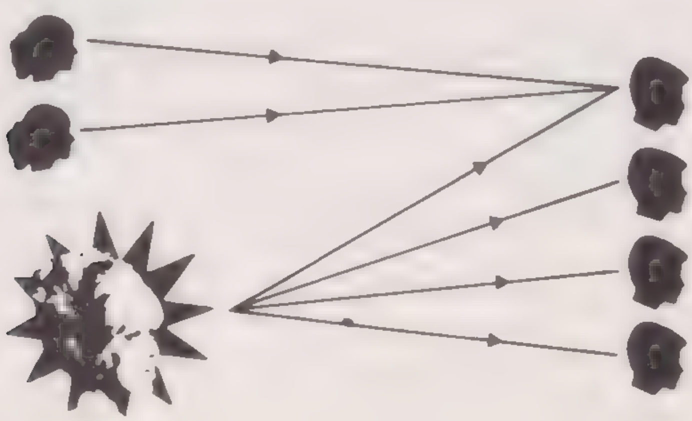
Each person has just one birthday, but many different people have the same birthday.



Example 7

MANY MANY MAPPING (NOT A FUNCTION)

The mapping from the set of all women who are or have been married to the set of all men, obtained by mapping a woman to her husband or ex-husband is MANY MANY, because some people get divorced and re-married several times, and their partner could have been married before. (The fact that this is not true of everyone is irrelevant. If it is true for some people, that will be sufficient to make the mapping many many.)

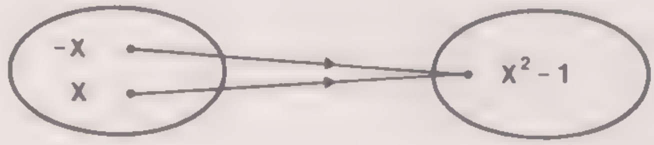


Example 8

The mapping defined by

$$f: x \longmapsto x^2 - 1 \qquad (x \in R)$$

is many one (a function). $x^2 - 1$ is defined as just one number when x is given a value. On the other hand, an image may correspond to more than one number; for example, $f(3) = 8$ and $f(-3) = 8$.



$f: x \longmapsto x^2 - 1$

Inverse Functions

We previously posed the question: "When is the reverse mapping a function?". We can now answer that question quite simply:

> The reverse mapping is a function if the original mapping is ONE–ONE or ONE–MANY. In both of these cases we can call the reverse mapping the reverse function.

The One–One Case

ONE–ONE functions are interesting, for if f maps A to B, and g is the reverse function of f: then f takes an element, a, in A to its image in B, and g brings this image back to a, (and to a only).

In other words

$$g(f(a)) = a \quad \text{for all } a \in A$$

Notice that this statement would also be true if f were a one–many mapping. But there is a difference; for one–one mappings we can *also* state

$$f(g(b)) = b \quad \text{for all } b \in f(A)$$

We now adopt the following definition:

> If f is a ONE–ONE function from A to B, and if $f(A) = B$ (i.e. the codomain of f is equal to the set of all images), then the function g from B to A where $g(f(a)) = a \ (a \in A)$ is called the INVERSE FUNCTION to f.

The condition which we impose on the codomain of f, namely that $f(A) = B$, is included to maintain the symmetry; so that f maps A to B, and g maps B to A. The condition $f(A) = B$ is equivalent to saying that B is the smallest set that will do the codomain's job; in other words, we do not want any odd elements like the unfortunate colour crimson in our "Colour of eyes" example.

For one–one functions whose domain and codomain are R, or subsets of R, we can often calculate inverses by algebraic manipulation.

Example 9

Suppose that we wanted to calculate the inverse function of f where

$$f : x \longmapsto 3x + 2 \quad (x \in R)$$

If we put

$$y = f(x)$$

then

$$y = 3x + 2$$

and we can calculate y if we are given x.

The inverse function will enable us to calculate x if we are given y, and we can find this inverse g by rearranging the equation $y = 3x + 2$ to give an equation of the form

$$x = \text{something involving } y \text{ (and not } x) = g(y)$$

If we do this we get

$$x = \frac{y-2}{3}$$

and so

$$g(y) = \frac{y-2}{3}$$

and therefore g is the mapping

$$g: y \longmapsto \frac{y-2}{3} \quad (y \in R)$$

We could of course rewrite this in the equivalent form

$$g: x \longmapsto \frac{x-2}{3} \quad (x \in R)$$

Exercise 1

Determine the inverse function of

$$f: x \longmapsto 4 - \frac{3}{x} \quad (x \in R^{+})$$

(Do not forget the domain of the inverse function.)

Exercise 2

If g is the inverse function of the one–one function f, is it true that

(i) f is the inverse function of g?
(ii) $g \circ f = f \circ g$?
(iii) $g(x) = \dfrac{1}{f(x)}$?

Inverses of Composite Functions

We have seen that the problem of finding an inverse function is essentially one of unravelling the original function. For functions from R to R this usually means unravelling a calculation. In some books this procedure goes under the name of "changing the subject of the formula".

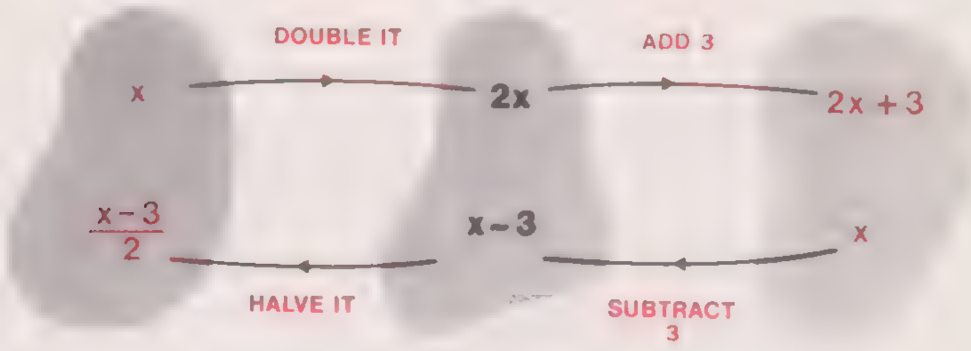
It is sometimes useful, when finding inverses of relatively simple functions, to decompose a function into more elementary ones.

Example 10

The function

$$f: x \longmapsto 2x + 3 \quad (x \in R)$$

has two components — "double it" and "add 3". If we want to invert this function we must unravel the calculation; "subtract 3" and "halve it".

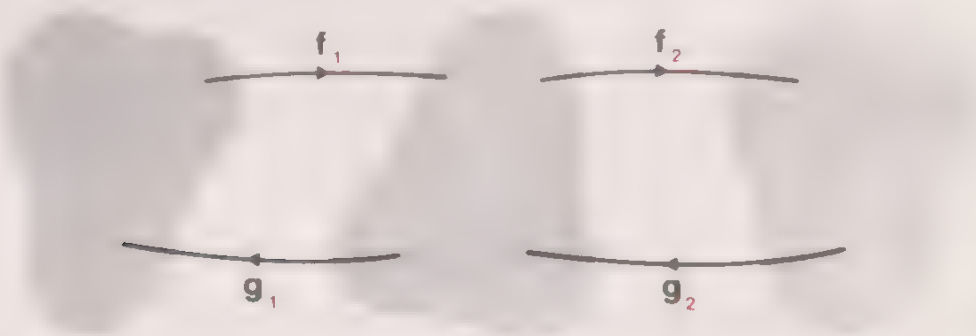


The inverse function is

$$g: x \longmapsto \frac{x - 3}{2} \quad (x \in R)$$

In general, if f_1 and f_2 are one–one functions and have inverses g_1 and g_2 then

$$\boxed{\text{the inverse of } f_2 \circ f_1 \text{ is } g_1 \circ g_2}$$



Note the order in which the inverses are combined — when inverting we have to invert the last step first. This is just like many operations in everyday life. When mending a puncture in a bicycle tyre, one removes the tyre first and then the inner tube. To invert this operation when the job is done, one replaces the inner tube first and then the tyre.

Exercise 3

The one–one function f where

$$f: x \longmapsto 3x^2 + 2 \quad (x \in R^+)$$

maps an element x to an element y, where

$$y = 3x^2 + 2$$

The inverse function, g, will map y back to x. By changing the subject of the formula, i.e. expressing x in terms of y, find a formula for g.

3.6 Additional Exercises

Exercise 1

Classify the following mappings as

$\boxed{\text{one one}}$ or $\boxed{\text{many one}}$ or $\boxed{\text{one many}}$ or $\boxed{\text{many many}}$

(i) $x \longmapsto |x^3 + 1|$ $(x \in R)$

(ii) $x \longmapsto \sin x + \cos x$ $(x \in R)$

(iii) $x \longmapsto (|x| - 1)^2$ $(x \in R)$

(iv) $x \longmapsto \{x, -x\}$ $(x \in R^+)$

(v) $x \longmapsto \{\sqrt{(1 + 3x^2)}, -\sqrt{(1 + 3x^2)}\}$ $(x \in R)$

Exercise 2

If f and g are functions:

$$f: x \longmapsto x^2 + 1 \quad (x \in R)$$
$$g: x \longmapsto 2x - 3 \quad (x \in R)$$

complete the following calculation of the function $g \circ f$.

The image of R under f is the set of real numbers greater than or equal to 1, and this is contained in the domain of g. The combination $g \circ f$ is therefore possible.

We can specify $g \circ f$ by giving a formula for $g(f(x))$ and stating the domain.

(i) The domain of f is?
Therefore the domain of $g \circ f$ is?

(ii) Let $y = f(x)$, so that $y = ?$

(iii) Let $z = g(y)$, so that $z = ?$

(iv) Substituting (ii) into (iii) gives $z = ?$

(v) $g \circ f$ is the function defined by?

Exercise 3

What can you say about the graph of a one–one function, assuming that it is an unbroken curve? (HINT: One approach to this might be to sketch a few graphs, decide whether or not they are one–one, and then try to generalize.)

Exercise 4

Calculate reverse mappings or inverse functions for the functions defined as follows:

(i) $f: x \longmapsto 7x - 1 \quad (x \in R)$
(ii) $f: x \longmapsto 4x^2 + 3 \quad (x \in R)$

3.7 Answers to Exercises

Section 3.1

Exercise 1

(i) Many–one, e.g. $1 \longmapsto 5$ and $-1 \longmapsto 5$.
(ii) One–one.
(iii) Many–one, e.g. $0 \longmapsto 0$, $\pi \longmapsto 0$, etc.

Section 3.2

Exercise 1

If you look at the definitions, we do our arithmetic in the codomain. Therefore, only the codomain need be R. The only restriction on the domain is that it should be the same for both f and g. One example, among many possibilities would be

$f: x \longmapsto$ mark for first assignment $\quad$ ($x \in$ set
$g: x \longmapsto$ mark for second assignment $\quad$ of students)

Exercise 2

(i) $g + f: x \longmapsto 6x + 6x^2 \quad (x \in [-1, 1])$
(ii) $g \div f: x \longmapsto 1/x \quad (x \in [-1, 1] \text{ and } x \neq 0)$
(iii) $f \div g: x \longmapsto x \quad (x \in [-1, 1] \text{ and } x \neq 0)$
(Although $f(x)/g(x) = x$, this is only true where we can actually perform the division. Therefore mathematically, and by definition, we must exclude $x = 0$. This is a substantial point.)
(iv) $f \times g: x \longmapsto 36x^3 \quad (x \in [-1, 1])$

Section 3.3

Exercise 1

(i) (a) $f \circ g: x \longmapsto x^2 - 1 \quad (x \in R)$
(b) $g \circ f: x \longmapsto (x - 1)^2 \quad (x \in R)$
(ii) It ought to be $g \circ f$, but in practice you would often get a different result going to German via French rather than directly from English.

Exercise 2

(i) NO. We can form $g \circ f$ only if the set of all images of the domain of f is a subset of (or equal to) the domain of g. Consider, for example,

f: person $\longmapsto$ colour of his eyes (domain the set of all people)

$g: x \longmapsto x^2 \quad (x \in R)$

Then we cannot form $g \circ f$, since the square of a colour is not defined.
(ii) NO. We now have restrictions on the set of images under g and on the domain of f. Consider, for example,

$f: x \longmapsto \sqrt{x} \quad (x \in R^+)$

$g: x \longmapsto x + 3 \quad (x \in R)$

Then

$g \circ f: x \longmapsto \sqrt{x} + 3 \quad (x \in R^+)$ is satisfactory

but

$f \circ g: x \longmapsto \sqrt{(x + 3)}$ is undefined for $x < -3$.

Section 3.4

Exercise 1

(i) $x \longmapsto x - \theta \qquad x \longmapsto \cos x \qquad x \longmapsto \dfrac{x}{2}$

$$x \longmapsto x - \theta \longmapsto \cos(x - \theta) \longmapsto \frac{\cos(x - \theta)}{2}$$

(ii)

x

↓

Subtract θ from the number

↓

Take the cosine of the number

↓

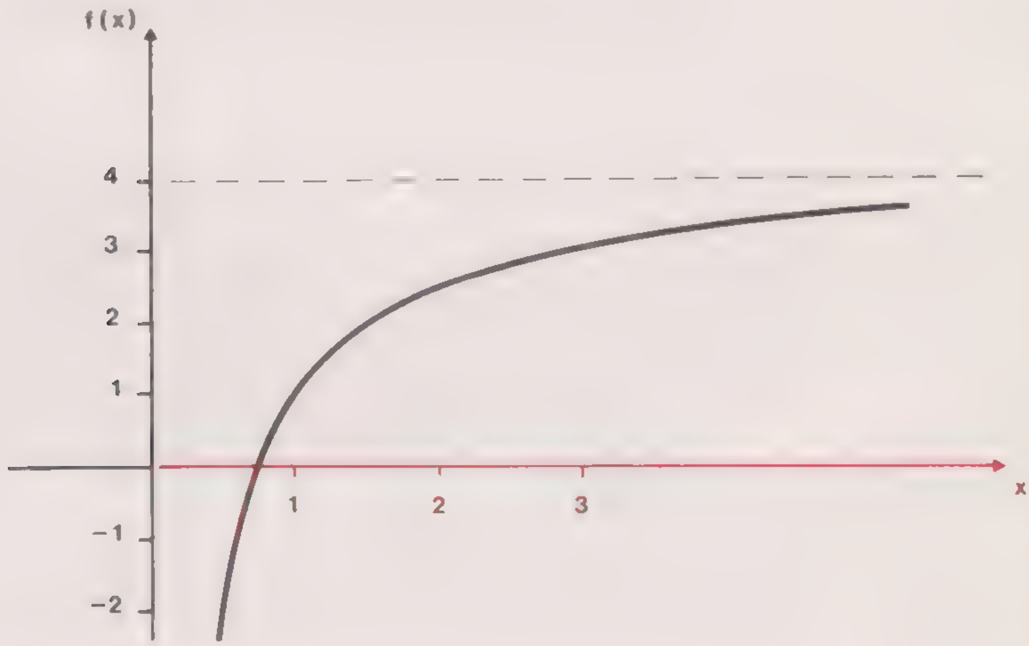
Halve the number

Section 3.5

Exercise 1

$$g : x \longmapsto \frac{3}{4 - x} \quad (x \in \text{the set of real numbers less than 4})$$

Remember that the domain is $f(R^+)$. Since x is positive, $4 - \dfrac{3}{x}$ is always less than 4. The graph of f shows quite clearly that the set of images is the set of all real numbers less than 4.



Exercise 2

(i) YES. We have actually defined the inverse function of f only when f is one–one. As we have hinted in the text, we could have defined

an inverse function of a one–many mapping f; then of course, f is not a function.

(ii) Both $g \circ f$ and $f \circ g$ are given by the formula $x \longmapsto x$. But the domain of $g \circ f$ is that of f, and the domain of $f \circ g$ is that of g. So $f \circ g = g \circ f$ only if f and g have the same domain.

For instance, in the previous exercise you can check that the formula for both $g \circ f$ and $f \circ g$ is $x \longmapsto x$. But the domain of $g \circ f$ is the domain of f, i.e. R^+, whereas the domain of $f \circ g$ is the domain of g, i.e. the set of real numbers less than 4.

Example 9 provides a case where $g \circ f = f \circ g$.

(iii) NO. If $f: x \longmapsto x$, then $g: x \longmapsto x$, so $g(x) \neq \dfrac{1}{x}$ unless the domain of f is $\{1\}$.

Exercise 3

Rearranging the equation we get

$$x = \pm\sqrt{\left(\frac{y-2}{3}\right)}$$

Since f is one–one we must choose the signs correctly and get a single value for x in terms of y. Since $x \in R^+$, we choose the positive sign and so

$$x = \sqrt{\left(\frac{y-2}{3}\right)}$$

or

$$g: y \longmapsto \sqrt{\left(\frac{y-2}{3}\right)}$$

To find the domain of g, we require $f(R^+)$. If $x \in R^+$ then $f(x)$ is always greater than 2. Thus, we finally get

$$g: x \longmapsto \sqrt{\left(\frac{x-2}{3}\right)} \quad (x \in \text{the set of real numbers greater than } 2)$$

(Notice that this is an example where the domain of f is critical. If f had domain R, it would not be one–one, and the reverse mapping (*not* inverse) would not be a function.)

Section 3.6

Exercise 1

(i) Many-one, e.g. $1 \longmapsto 2$ and $-\sqrt[3]{3} \longmapsto 2$

(ii) Many-one, e.g. $0 \longmapsto 1$ and $\frac{\pi}{2} \longmapsto 1$

(iii) Many-one, e.g. $-1 \longmapsto 0$ and $1 \longmapsto 0$

(iv) One many. (Had the domain been R instead of R^+, the mapping would have been many–many.)

(v) Many–many, e.g. $1 \longmapsto \{2, -2\}$, $-1 \longmapsto \{2, -2\}$.

Exercise 2

(i) R, R

(ii) $y = x^2 + 1$

(iii) $z = 2y - 3$

(iv) $z = 2(x^2 + 1) - 3$

(v) $x \longmapsto 2x^2 - 1 \quad (x \in R)$

Exercise 3

The graph cannot "bend back" on itself. $f(x)$ must either increase steadily or decrease steadily as x increases.

Exercise 4

(i) $g: x \longmapsto \dfrac{x+1}{7} \qquad (x \in R)$

(ii) $g: x \longmapsto \left\{\sqrt{\left(\dfrac{x-3}{4}\right)}, -\sqrt{\left(\dfrac{x-3}{4}\right)}\right\}$ ($x \in$ the set of real numbers greater than or equal to 3)

CHAPTER 4 LIMITS

4.0 Introduction

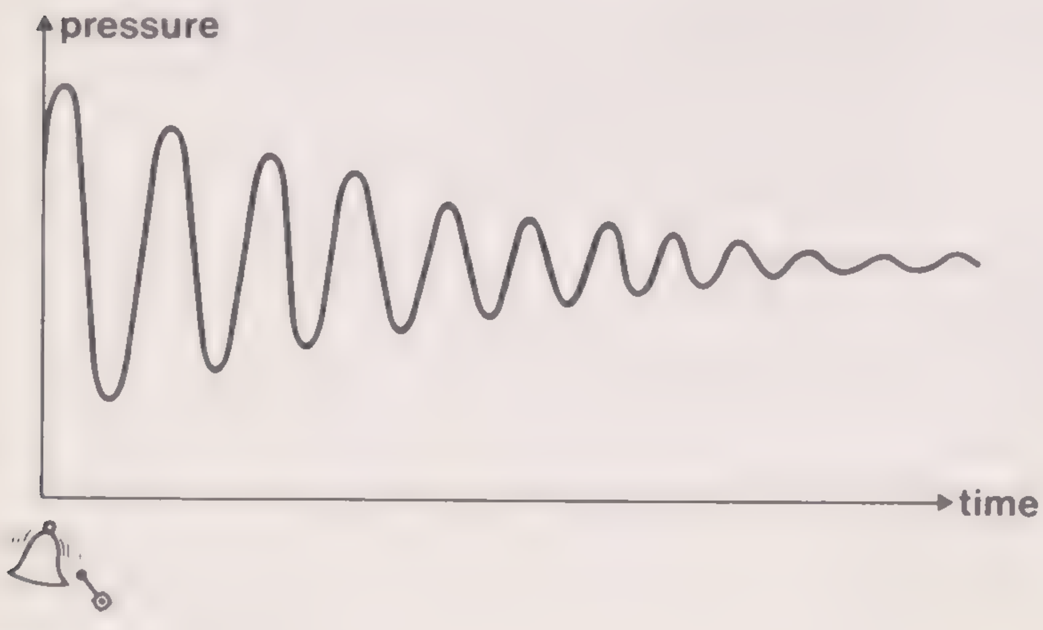
In Chapter 2 we introduced very intuitively the idea of a limit of a sequence.

In this chapter we are going to extend the idea of a limit to *real functions*. These are functions whose domain and codomain are R or subsets of R. First, we shall consider the limit of a function of a real variable x as x becomes very large. Secondly, we shall consider the idea of the limit of a real function at any point.

This second idea will enable us to lead on to the concept of the continuity of a function. However, we shall continue with our intuitive approach throughout this chapter, leaving a more rigorous treatment until Chapter 6.

4.1 Limits of Real Functions

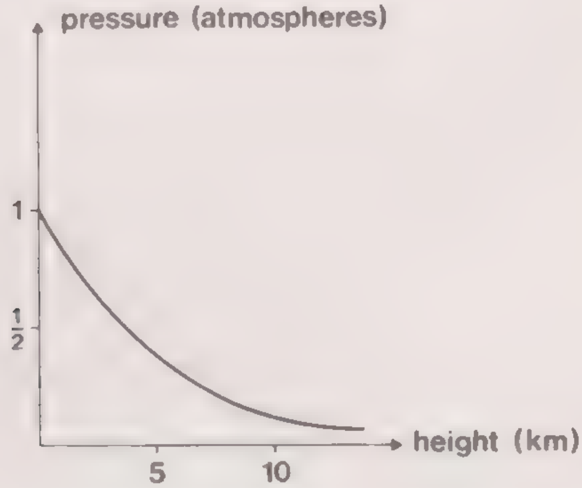
Although the notion of a limit takes its simplest form when applied to sequences, it is by no means restricted to sequences in its application. For example, when a bell is struck, the pressure in the air nearby will vary with time roughly as shown below.



The amplitude of the oscillations decreases slowly with time, so that a long time after the bell is struck the pressure is approximately constant and equal to the atmospheric pressure.

This graph is qualitatively similar to the graph of a convergent sequence, for example part (vi) of Exercise 2.3.1 (page 52), and so we may expect to be able to use the ideas of convergence and of limits here too. In both cases we are dealing with a function: in the bell example because the air pressure depends on the time elapsed since the bell was struck, and in the sequence example because the kth term of the sequence depends on k. The main difference between these two functions is that in the bell example the domain is R^+ whereas for an infinite sequence the domain is Z^+.

Despite this difference, the intuitive definition of a limit can be carried over quite easily from the case of sequences to the case of functions with domain R^+ and codomain R or subsets of R. As another example, here is a graph showing the dependence of atmospheric pressure (at some particular instant of time) on height above the earth's surface.



When the height is very large, the atmospheric pressure is very small.

The similarity of this statement to the Intuitive Definition of a limit of an infinite sequence (page 49) suggests that the concept of a limit will also be useful here. Accordingly we adopt the following:

> **Intuitive Definition of a Limit**
>
> If f is a real function and L is a number, then "L is the limit of f for large numbers in its domain" is equivalent to the statement "whenever x is very large, $f(x)$ is a very good approximation to L".

We use the notation $\lim\limits_{x \text{ large}} f(x)$ for the limit of f for large numbers in its domain.* Thus, if f is the function mapping elapsed time to pressure in the first graph in this section, then $\lim\limits_{x \text{ large}} f(x)$ is the mean atmospheric pressure; and if g is the function mapping height to pressure in the second graph, then $\lim\limits_{x \text{ large}} g(x)$ is 0. The same notation can also be used for limits of sequences; thus the limit of the sequence $1, \frac{1}{2}, \frac{1}{3}, \ldots$ may be written $\lim\limits_{k \text{ large}} \dfrac{1}{k}$.

Exercise 1

The following functions all have domain R^+. Which ones have a limit for large numbers in their domain, and what are the limits?

(i) $t \longmapsto 4$

(ii) $t \longmapsto t^2$

(iii) $t \longmapsto \dfrac{1}{t}$

(iv) $t \longmapsto \sin t$

(v) $t \longmapsto \dfrac{\sin t}{t}$

Limit Near a Point

There is another way of applying the concept of a limit to functions. This time the analogy with limits of sequences, though still present, is not quite as close. Let us consider just what we mean by the "rate of change" of the images under a function. For example, the position of a car moving north on a straight road out of London may be specified fairly closely by giving its distance from its initial position (Westminster say). This distance depends on time: let us denote it by $f(t)$, where t is the time that has elapsed since the car left London at (say) noon. If we know the function f, how do we calculate the velocity**, that is, the rate of change of $f(t)$?

* Any letter could be used here in place of x, e.g. $\lim\limits_{t \text{ large}} f(t)$. Notations such as $\lim\limits_{x \to \infty} f(x)$ are very commonly used to mean exactly the same thing, but we prefer to avoid the symbol ∞ at this stage, because it is dangerous unless fully understood.

** "Velocity" means speed in a known direction. Here only two directions are possible: away from London and towards it. We distinguish them by giving the velocity a positive sign for motion away from London, and a negative sign for motion towards it.

It is easy enough to calculate the *average* velocity of the car over some specified time interval; it is the distance travelled in that time interval divided by the duration of the interval. For example, if the car travels 11 miles in 10 minutes, then its average velocity over this time interval is $\frac{11}{\frac{1}{6}} = 66$ mile/h (since 10 minutes $= \frac{1}{6}$ hour). This rule for calculating average velocities can be written as a formula:

$$\text{average velocity} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{f(t_2) - f(t_1)}{t_2 - t_1}$$ Equation (1)

$$= w(t_1, t_2), \text{ say}$$

where t_1 and t_2 are the times since noon at the beginning and end of the time interval respectively, and $x_1 = f(t_1)$, $x_2 = f(t_2)$ are the distances from London at the beginning and end of the time interval respectively. This formula defines $w(t_1, t_2)$ for $t_1 < t_2$, and also for $t_1 > t_2$, but not for $t_1 = t_2$.

In many cases, however, the really important velocity is not the average velocity but the *instantaneous* velocity.

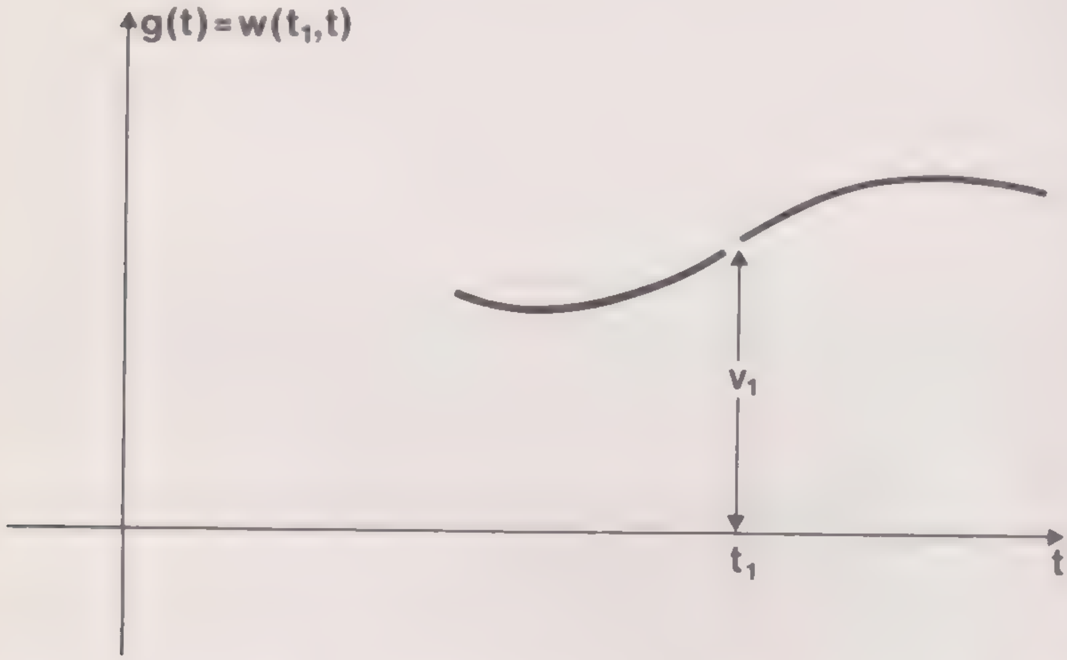
Suppose for example that the car had a collision; then the velocity of vital importance to the occupants of the car is its velocity at the instant of impact, not its average over the previous ten minutes, or even the previous ten seconds, during which the driver may have been braking violently. The obvious way to try to get instantaneous velocities from Equation (1) is to consider a time interval of vanishing duration, that is, to set $t_2 = t_1$. Since the distance travelled by the car in this zero time interval is zero, the fraction in Equation (1) now takes the form $\frac{0}{0}$; this expression, however, is nonsense because there is no mathematically consistent way of defining division by zero. By making $(t_2 - t_1)$ small we can calculate the average velocity over as short a time interval as we please, but as soon as we try to catch the instantaneous velocity by making this time interval exactly zero, the quantity we are looking for slips from our grasp.

In order to catch this elusive fish (assuming, of course, that it really exists), a more sophisticated technique is necessary, in which we deduce the instantaneous velocity from the average velocities over very short time intervals. To construct a definition of instantaneous velocity in terms of the average velocities, we make use of the concept of a limit. It is reasonable to assume that the instantaneous velocity does not change very rapidly or fluctuate wildly, and hence that the average velocity, over a very short time interval which includes the instant t_1, will be a very good approximation to the instantaneous velocity. For simplicity, let us look for a definition of the instantaneous velocity, v_1, at a particular instant, t_1,

in terms of the average velocities over time intervals beginning or ending at t_1.

In other words, if we define a function g by:

$$g : t \longmapsto w(t_1, t) \quad (t \in \text{domain of } f \text{ and } t \neq t_1)$$



then $g(t)$ is a very good approximation to v_1 whenever t is very close to t_1, but $g(t_1)$ does not exist. This is just like the Intuitive Definition on page 83 for the limit of a function for very large values in its domain, except that here we are concerned with values very close to t_1 instead of very large values. We have thus arrived at a new type of limit, whose intuitive definition may be stated as follows:

> **Intuitive Definition of a Limit**
>
> If g is a real function and a and L are real numbers, "L is the limit of g near a" is equivalent to the statement "if x is very close to a, but not equal to it, then $g(x)$ is very close to L".

The notation we shall use for the limit of g near a is

$$\lim_{x \rightsquigarrow a} g(x)$$

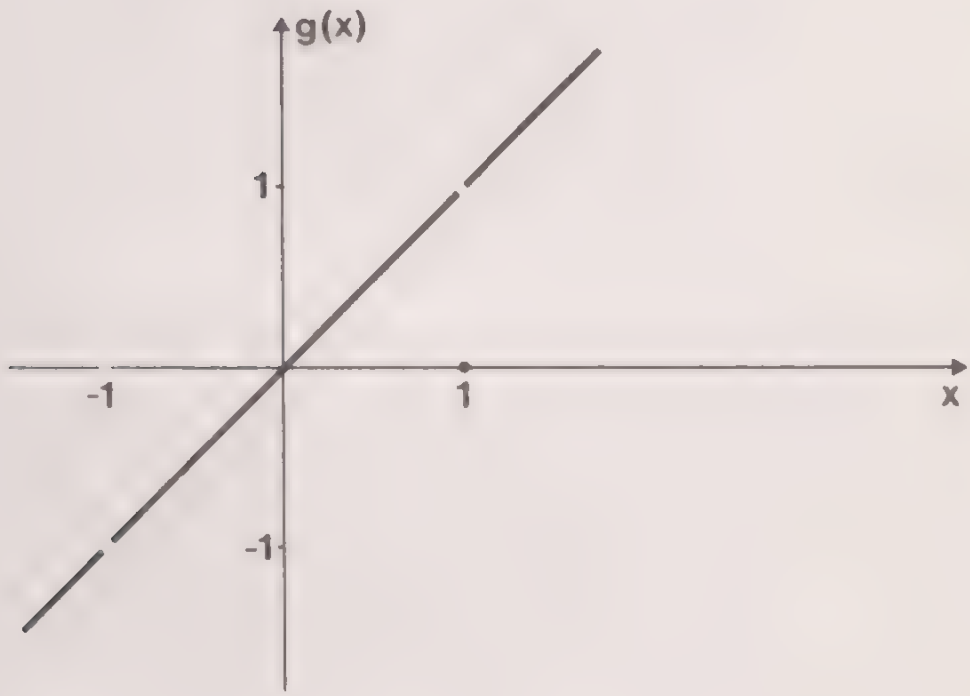
The notation

$$\lim_{x \to a} g(x)$$

is more common but we are already using the straight arrow for mappings.

The limit of g near a is often described as "the limit of $g(x)$ as x approaches (or tends to) a" since it is often convenient to think of the limit in terms of motion: if the value of x changes with time, moving steadily towards a (without ever actually getting there), the value of $g(x)$ changes steadily and eventually gets as close as we please to $\lim_{x \to a} g(x)$.

In this definition you should notice that a itself does not have to belong to the domain of g: the definition is concerned with values of x that are very close to a, but not with $x = a$. If a does belong to the domain of g, the value of $g(a)$ does not affect the value of $\lim_{x \to a} g(x)$ and the two numbers may be either the same or different. If a does not belong to the domain of g then $g(a)$ does not exist, but even so $\lim_{x \to a} g(x)$ may still exist. In the following example the limits for x near 1 and -1 have just the same values as if the graph were an unbroken straight line:



The above diagram is a graph of the function g defined by

$$g: x \longmapsto \begin{cases} x & \text{if } x \neq 1 \\ 0 & \text{if } x = 1 \end{cases} \quad (x \in R \text{ and } x \neq -1)$$

In this graph

$$\lim_{x \to 0} g(x) = 0 = g(0)$$

but

$$\lim_{x \to 1} g(x) = 1 \neq g(1)$$

and

$$\lim_{x \to -1} g(x) = -1, \text{ but } g(-1) \text{ does not exist.}$$

(Cf. our velocity example to which we referred above: it is just because $g(t_1)$ is not defined that we use $\lim_{x \to t_1} g(x)$ as our definition of v_1.)

Exercise 2

Draw the graphs of the following functions and determine their limits near 1.

(i) g_1, where $g_1 : x \longmapsto x + 1 \qquad (x \in R)$

(ii) g_2, where $g_2 : \begin{cases} x \longmapsto x + 1 & \text{if } x \neq 1 \\ x \longmapsto 0 & \text{if } x = 1 \end{cases} \qquad (x \in R)$

(iii) g_3, where $g_3 : x \longmapsto \dfrac{(x^2 - 1)}{x - 1} \qquad (x \in R \text{ and } x \neq 1)$

Exercise 3

Is the following statement true or false? If true, give a demonstration* or proof: if false, give a counter-example.

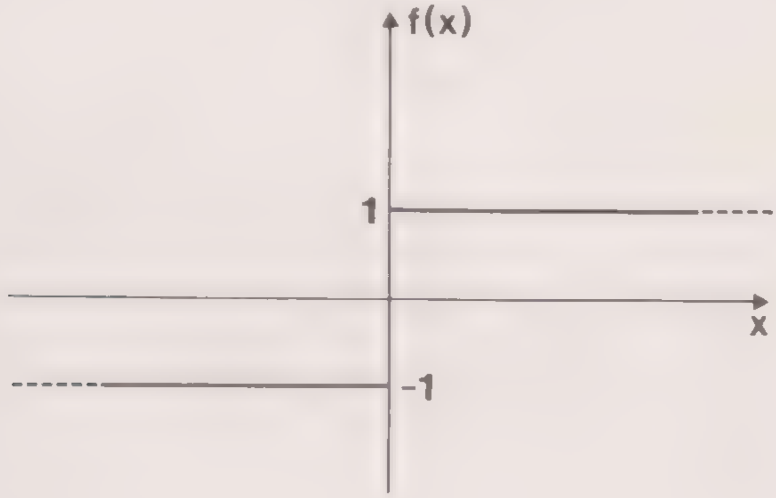
"If $\lim_{x \to a} f(x) = L$, where a and L are real numbers and f is a real function, and $x_1, x_2, x_3, \ldots$ is a sequence with limit a, none of whose elements is equal to a, then the limit of the sequence $f(x_1), f(x_2), \ldots$ is L."

* By "demonstration" we mean an argument that is not a (rigorous) proof, for example an argument based on a diagram or on an "intuitive definition".

4.2 Continuity

The graphs of some real functions can be drawn without lifting pen from paper, while for others this is not possible. The former are described as continuous functions, and we can express this idea of continuity in mathematical terms using the concept of the limit of a function. We begin by considering some examples.

Example 1

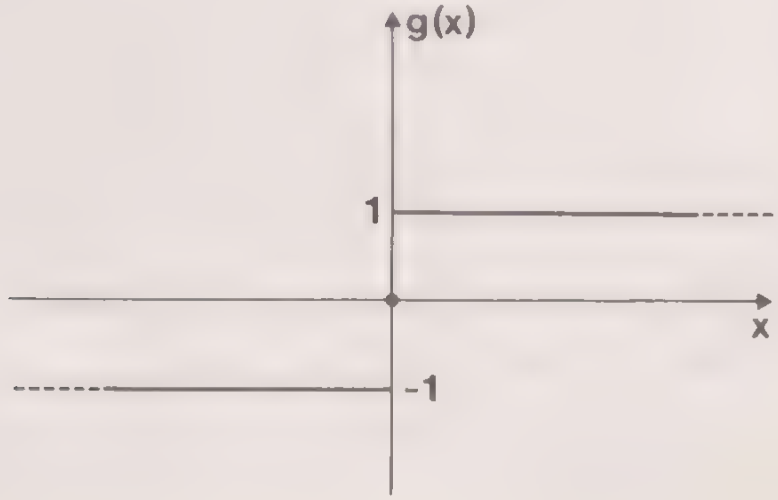


This is the graph of the function:

$$f : x \longmapsto \begin{cases} +1 \text{ if } x > 0 \\ -1 \text{ if } x < 0 \end{cases}$$

with domain R (excluding zero).

Example 2

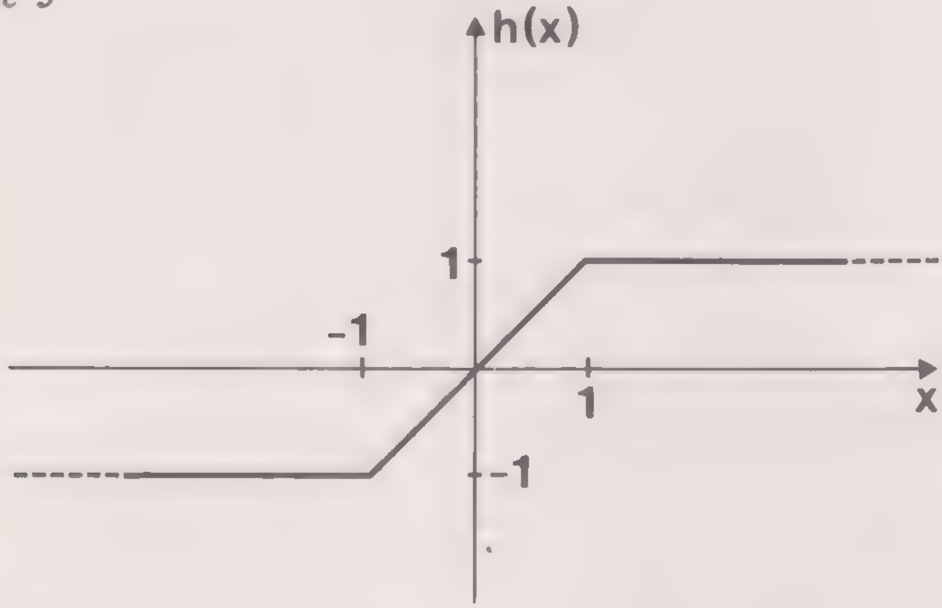


This is the graph of the function:

$$g : x \longmapsto \begin{cases} +1 \text{ if } x > 0 \\ 0 \text{ if } x = 0 \\ -1 \text{ if } x < 0 \end{cases} \quad (x \in R)$$

It is sometimes called the sign (not sine) function, since $g(x)$ has the sign, but not the magnitude, of x.

Example 3



This is the graph of the function:

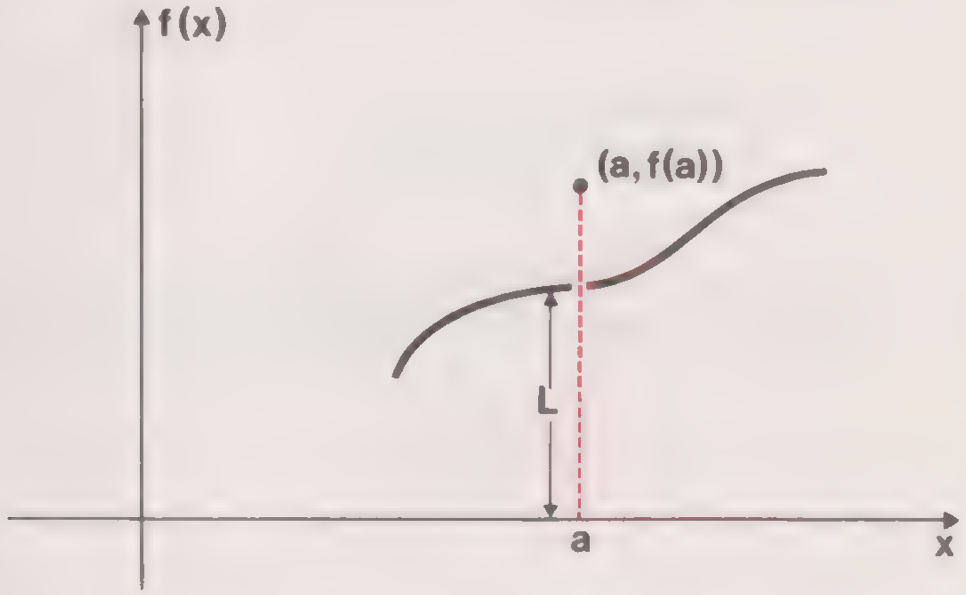
$$h : x \longmapsto \begin{cases} -1 \text{ if } x \leqslant -1 \\ x \text{ if } -1 < x < 1 \\ +1 \text{ if } x \geqslant 1 \end{cases} (x \in R)$$

The combined *gap* and *jump* in the graph in Example 1 has the effect that if x is very small indeed then a slight change in x can produce a change of magnitude of 2 in $f(x)$. In a similar way, the jumps in the graph in Example 2 have the effect that a slight change in x near the value 0 can produce a change of magnitude of 1 or 2 in $g(x)$.

In Example 3, the graph has neither a gap nor a jump. A small change in x necessarily produces only a small change in $h(x)$. The functions of

Examples 1 and 2 are said to be *discontinuous* at 0, but *continuous* elsewhere. The function of Example 3 is continuous everywhere in its domain.

For a precise definition of continuity, the concept of limit of a function at a point serves admirably. For any real function f, we have defined the limit of f near a to be a number L such that if x is very close, but not equal, to a, then $f(x)$ is very close to L. If this limit exists, therefore, and a is in the domain of f, then the only possible gap in the graph when x is close to a is a displaced point at $x = a$ itself; for example:

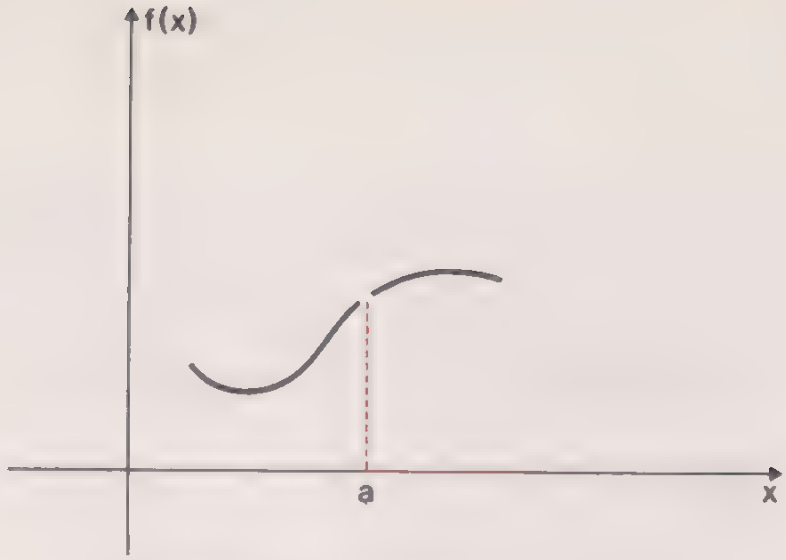


However, if the limit near a not only exists but is equal to $f(a)$, then the function has no gap or jump at a and may therefore be said to be continuous at a.

Accordingly we make the

> **Definition of Continuity**
>
> If f is a real function and a is an element of its domain, then "f is continuous at a" is equivalent to the statement "$\lim_{x \to a} f(x)$ exists and is equal to $f(a)$".

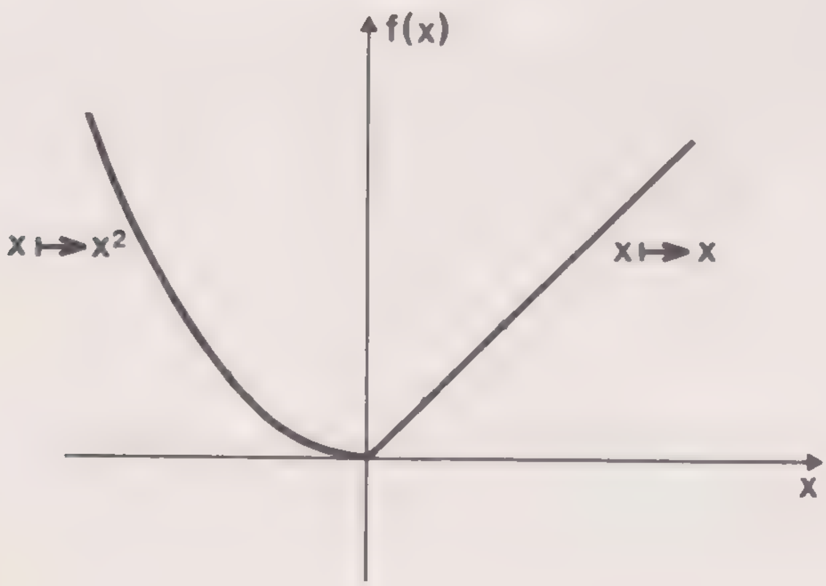
Notice that we now require a to belong to the domain, whereas in defining $\lim_{x \to a} f(x)$ we did not; by this definition, if $f(a)$ is undefined (i.e. if a does not belong to the domain), then f is not continuous at a. Thus the definition fits our intuitive ideas of continuity in this case too, since the graph must have a gap at a if $f(a)$ is undefined.



Example of a function f for which f(a) is not defined

The practical way of discovering whether a function is continuous or not is to sketch its graph. For example, here are graphs of a few functions:

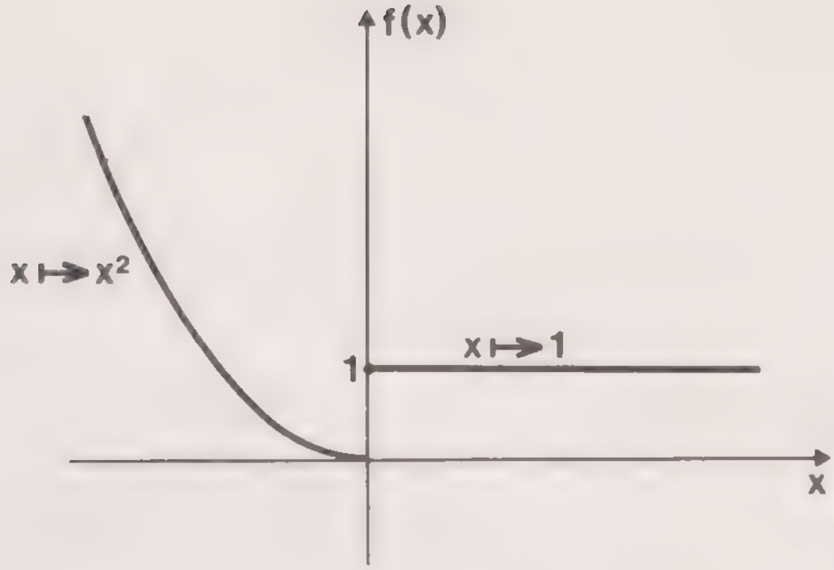
Example 4



$$f: x \longmapsto \begin{cases} x & \text{if } x \geqslant 0 \\ x^2 & \text{if } x < 0 \end{cases} \quad (x \in R)$$

is continuous throughout its domain.

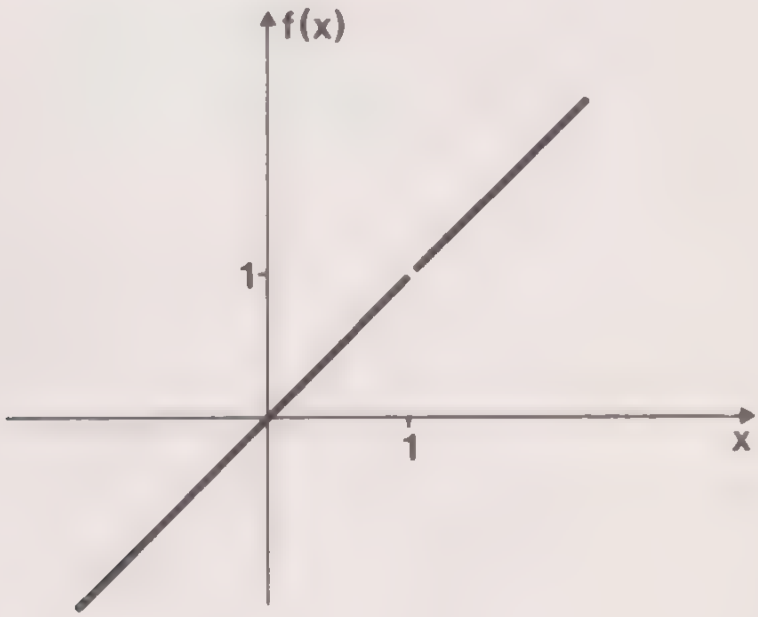
Example 5



$$f : x \longmapsto \begin{cases} 1 & \text{if } x \geqslant 0 \\ x^2 & \text{if } x < 0 \end{cases} \quad (x \in R)$$

is discontinuous at 0 (because of the jump), but continuous elsewhere.

Example 6



$$f : x \longmapsto x \quad (x \in R \text{ and } x \neq 1)$$

is discontinuous at 1 (because of the gap in the domain), but continuous elsewhere.

Exercise 1

Which of the following functions are continuous at 0?

(i) h_1, where $h_1 : \begin{cases} x \longmapsto x & \text{if } x \geqslant 0 \\ x \longmapsto 0 & \text{if } x < 0 \end{cases}$ $(x \in R)$

(ii) h_2, where $h_2 : \begin{cases} x \longmapsto \dfrac{|x|}{x} & \text{if } x \neq 0 \\ x \longmapsto 1 & \text{if } x = 0 \end{cases}$ $(x \in R)$

(iii) h_3, where $h_3 : x \longmapsto \dfrac{1}{x^2}$ $(x \in R \text{ and } x \neq 0)$

(iv) h_4, where $h_4 : x \longmapsto 1$ $(x \in R \text{ and } x \neq 0)$

4.3 Additional Exercises

Exercise 1

If a and L are real numbers, and g is a real function then $L = \lim_{x \to a} g(x)$ implies which one of the following?

(i) Given any positive number ε, it is possible to find a positive number δ such that, for all x in $[a - \delta, a + \delta]$ we have $g(x) \in [L - \varepsilon, L + \varepsilon]$.
(ii) For each positive number ε, there exists a positive number δ such that the image under g of the set $\{x : 0 < |x - a| \leqslant \delta\}$ is a subset of $[L - \varepsilon, L + \varepsilon]$.
(iii) $L = g(a)$.

Exercise 2

State where (if anywhere) each of the following functions is discontinuous.

(i)

$$f : x \longmapsto \begin{cases} 1 & \text{if } x < 0 \\ x^2 + 1 & \text{if } x \geqslant 0 \end{cases} \qquad (x \in R)$$

(ii)

$$g : x \longmapsto \begin{cases} 1 & \text{if } x < 0 \\ x^2 + 1 & \text{if } x > 0 \end{cases} \qquad (x \in R \text{ excluding } 0)$$

(iii)

$$h : x \longmapsto \begin{cases} -(x^2 + 1) & \text{if } x < 0 \\ x^2 + 1 & \text{if } x \geqslant 0 \end{cases} \qquad (x \in R)$$

4.4 Answers to Exercises

Section 4.1

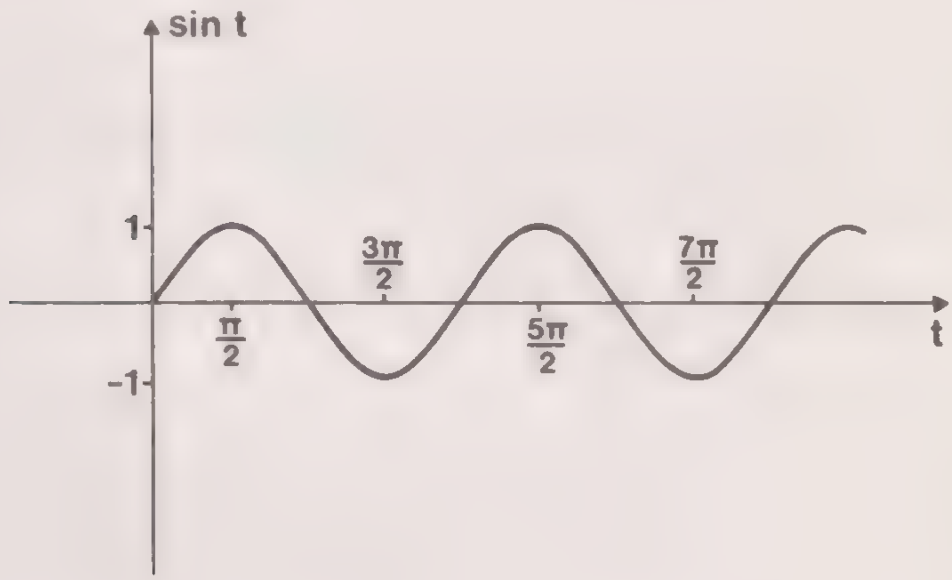
Exercise 1

(i) 4.

(ii) No limit.

(iii) 0; for if ε is any given positive number, then for all $t > \frac{1}{\varepsilon}$, $\frac{1}{t}$ lies in the interval $[0 - \varepsilon, 0 + \varepsilon]$.

(iv) No limit.

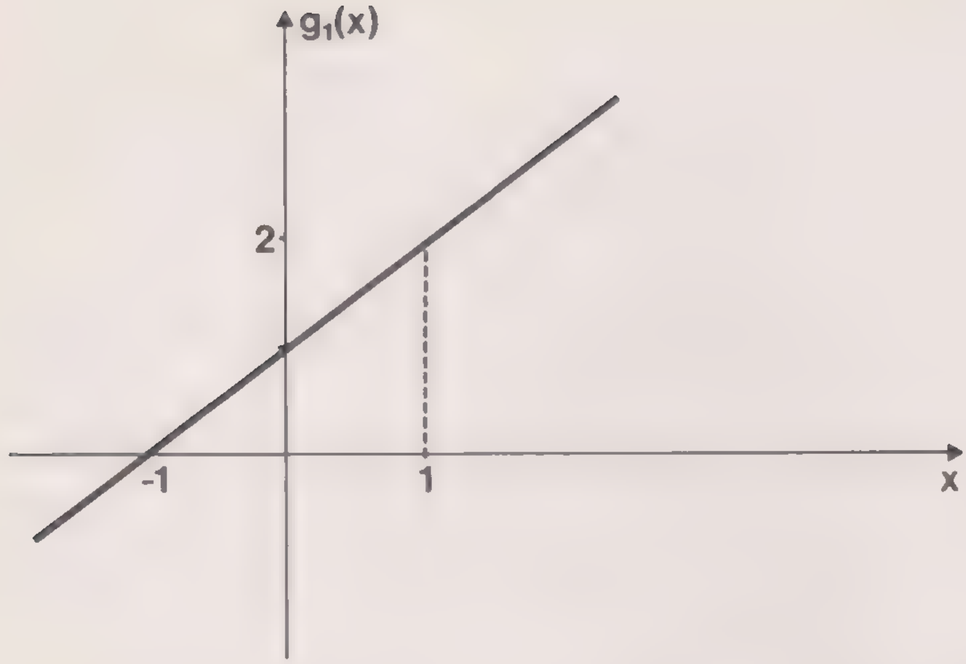


The graph continues to oscillate between 1 and -1 and thus, as with the sequence $1, -1, 1, -1, \ldots$ there is no limit.

(v) 0; for if ε is any given positive number, we can ensure that $\left|\frac{\sin t}{t}\right| < \varepsilon$ by choosing $t > \frac{1}{\varepsilon}$, since $|\sin t| \leqslant 1$.

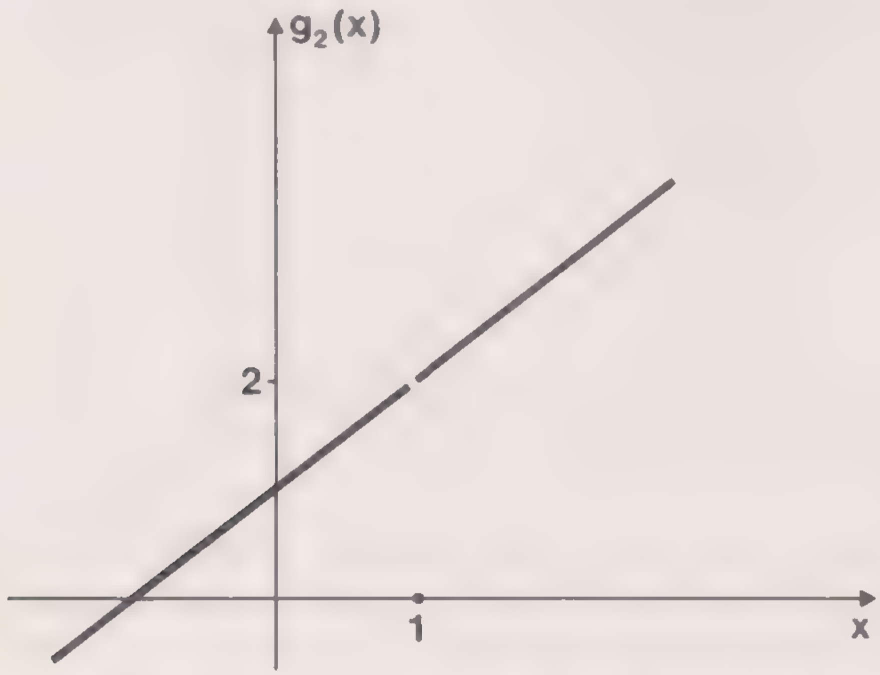
Exercise 2

(i)



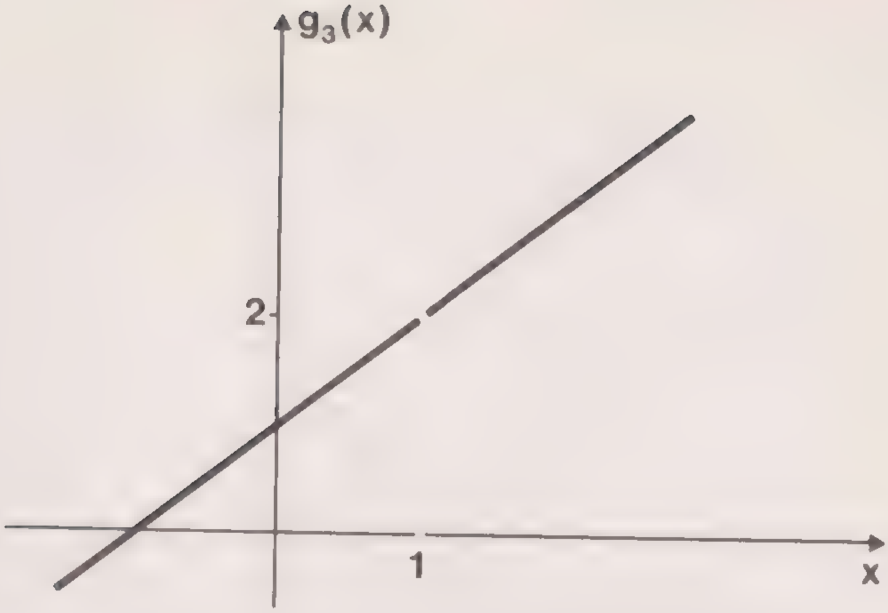
We see that $\lim_{x \to 1} g_1(x) = 2$. In this case $g_1(1)$ exists and also equals 2.

(ii)



Even though $g_2(1) = 0$ the limit is the same as for (i). In this case $g_2(1)$ exists (its value is 0) but is not the same as $\lim_{x \to 1} g_2(x)$, whose value is 2.

(iii)



In this case

$$g_3(x) = \frac{x^2 - 1}{x - 1} = \frac{(x-1)(x+1)}{x-1}$$

$$= x + 1$$

(The division by $x - 1$ is legitimate; $x - 1$ is never zero since 1 does not belong to the domain of g_3.) Although $g_3(1)$ is not defined, $\lim_{x \to 1} g_3(x)$ exists and is equal to 2.

Exercise 3

The statement is true.

We have two bits of information:

(i) $\lim_{k \text{ large}} x_k = a$,

(ii) $\lim_{x \to a} f(x) = L$.

Item (i) tells us that

if k is large, then x_k is close to a;

item (ii) tells us that

if x_k is close to a, then $f(x_k)$ is close to L.

Combining them tells us that

if k is large, then $f(x_k)$ is close to L,

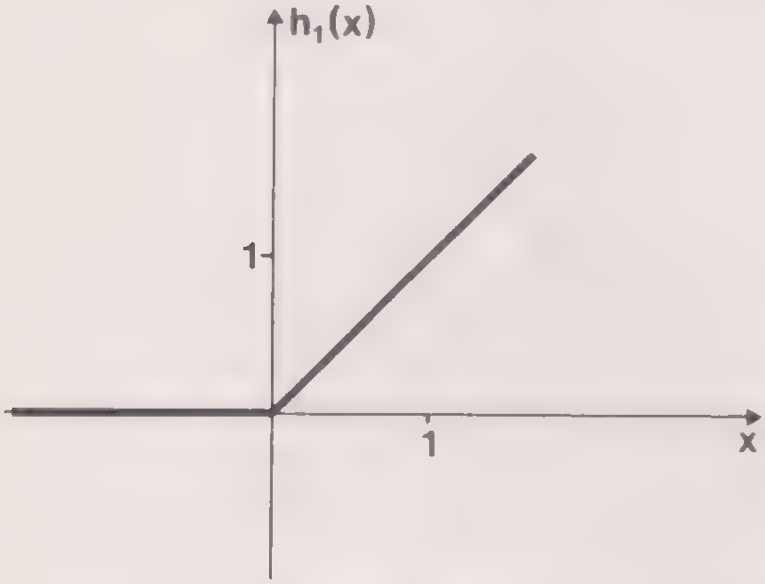
or in other words that

$$\lim_{k \text{ large}} f(x_k) = L.$$

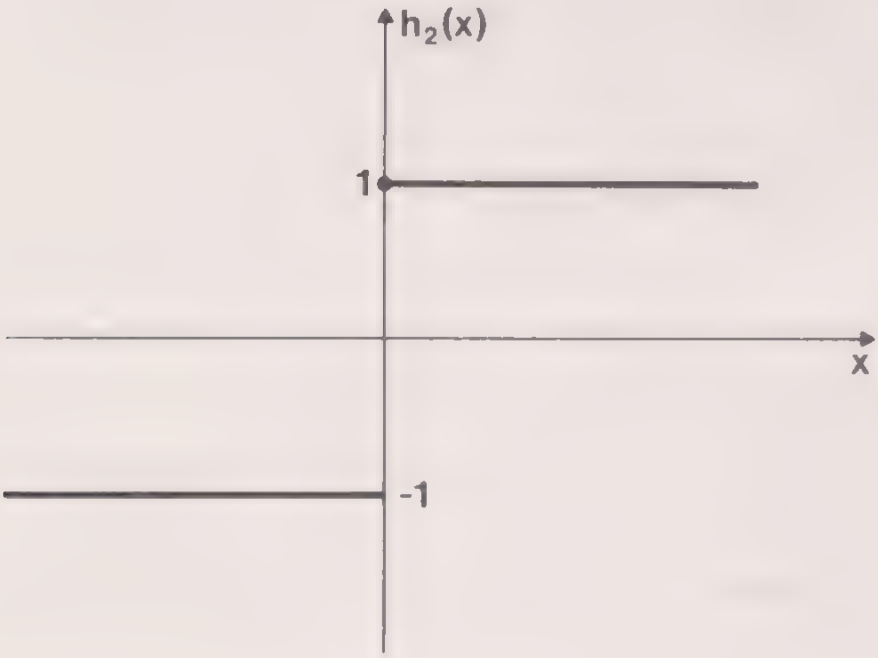
Section 4.2

Exercise 1

h_1 is the only function listed that is continuous at 0. Its graph is:



Since $|x| = x$ for $x > 0$ but $|x| = -x$ for $x < 0$, the graph of h_2 is:



which has a gap at $x = 0$, and so h_2 is not continuous at 0.

h_3 and h_4 are not continuous at 0 since the functions are not defined there.

Section 4.3

Exercise 1

Alternative (ii) is correct.

Alternative (iii) is wrong because $g(a)$ need not equal $\lim_{x \to a} g(x)$.

See (ii) of Exercise 4.1.2 for an example where they are different. The limit L as x tends to 1 is 2, but the image of 1 is zero.

Alternative (i) is wrong because it refers to "all x in $[a - \delta, a + \delta]$" when, in fact, the point $x = a$ should be excluded from the set of values of x considered.

Once again (ii) of Exercise 4.1.2 gives a counter-example. Here we have

$$g(x) = \begin{cases} x + 1 & \text{if } x \neq 1 \\ 0 & \text{if } x = 1 \end{cases} \quad (x \in R)$$

and so, if $x \in [1 - \delta, 1 + \delta]$, then $g(x) \in [2 - \delta, 2 + \delta]$ if $x \neq 1$, but $g(1) = 0$. It follows that, given any small positive ε, it is impossible to find a positive number δ such that, for all x in $[1 - \delta, 1 + \delta]$, $g(x) \in [L - \varepsilon, L + \varepsilon]$ as required by (i).

Alternative (ii) is correct: it differs from (i) only by excluding the point $x = a$ from the set of x-values considered. In fact this statement is the basis of the following more comprehensive definition of $\lim_{x \to a} g(x)$:

A limit of a function g near a point a is a number L such that for each positive number ε, however small, there is a positive number δ such that the set $\{x: 0 < |x - a| \leqslant \delta \text{ and } x \in \text{the domain of } g\}$ is nonempty, and its image under g is a subset of $[L - \varepsilon, L + \varepsilon]$.

Exercise 2

(i) f is continuous everywhere.
(ii) g is discontinuous at 0 because 0 is excluded from its domain.
(iii) h is discontinuous at 0 because there is a jump from $g(x) = -1$ to $g(x) = +1$ as x increases through the value 0.

CHAPTER 5 THE EXPONENTIAL FUNCTION

5.0 Introduction

In Chapter 4 we have considered briefly the concept of a limit. This is a powerful concept as we shall see in this chapter when we use it to define new numbers and new functions which cannot be defined using only the finite processes of ordinary arithmetic.

We introduce the exponential function by first looking at population growth. This leads us to consider the convergence of a particular sequence and to introduce the important number e and the function $x \longmapsto \exp(x)$ $(x \in R)$.

Next, we present the exponential theorem and natural logarithms, and we conclude by giving two formal proofs of results which we have used earlier in the chapter, but did not prove at the time because we did not want to interrupt the flow of the "story".

5.1 Population Growth

The significance of the exponential function is that it provides the simplest mathematical representation for growth processes and also for decay processes. An example is the growth of the world's population: the "population explosion". We can set up a mathematical model of this by denoting the time (i.e. the number of years that have elapsed since some designated initial instant) by t, the population at time t by $f(t)$ (i.e. f is the function that maps the time to the population at that time), and the birth and death rates per annum per head of population by b and d. We assume for simplicity that b and d are constants. As a first step towards determining how $f(t)$ depends on t, let us look at the population changes during a single year, lasting from, say, time t_0 to $t_0 + 1$. The calculation can be laid out in this way (the sign $\simeq$ means "approximately equals"):

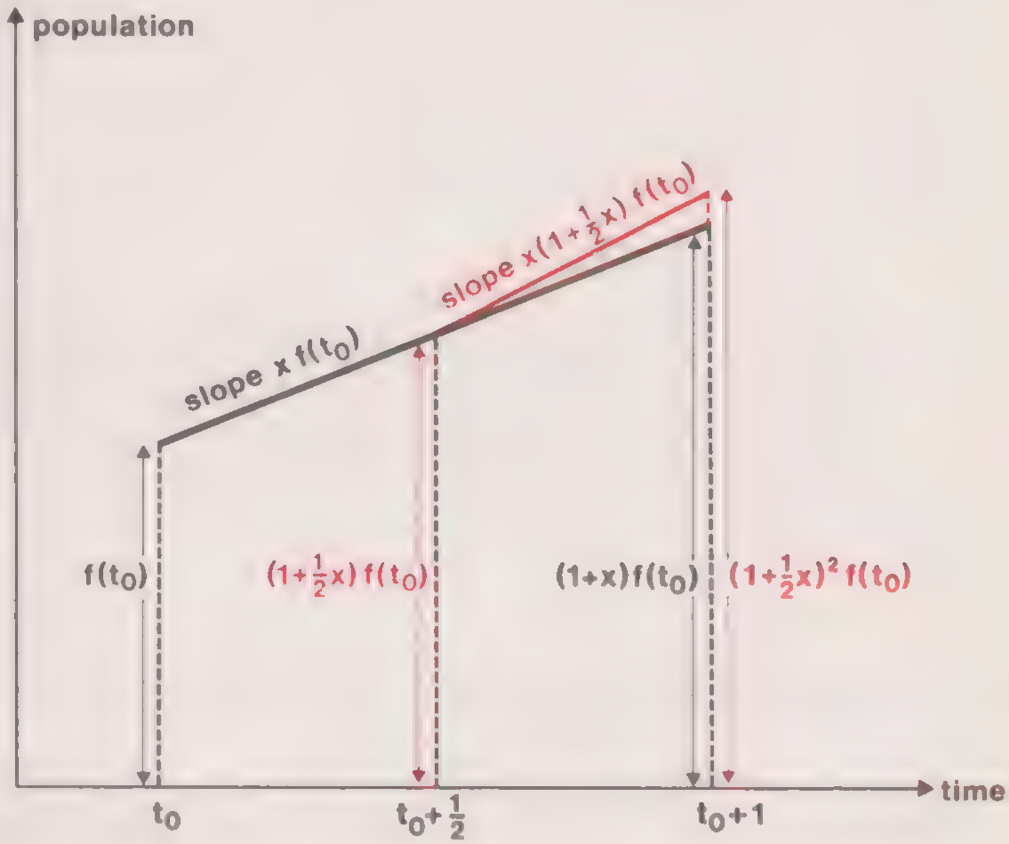
number alive at time t_0	$= f(t_0)$;
number born between t_0 and $t_0 + 1$	$\simeq bf(t_0)$;
number dying between t_0 and $t_0 + 1$	$\simeq df(t_0)$;
number alive at time $t_0 + 1$	$\simeq f(t_0) + bf(t_0) - df(t_0)$

i.e. $f(t_0 + 1) \simeq (1 + x)f(t_0)$ **Equation (1)**

where we define x, the net rate of population increase per annum per head of population, by

$$x = (b - d)$$

Can you see why we used the sign "$\simeq$" instead of "$=$"? The inaccuracy is that we have calculated the births and deaths as if the population had the constant value $f(t_0)$ throughout the year. For a more accurate calculation we should allow for the fact that $f(t)$ increases throughout the year, so that the population is greater in the second half year than in the first, and consequently (with constant birth and death rates) there are more births and deaths in the second half of the year than in the first. One way to do this is to consider the two halves of the year separately as shown below:



Black : This part is based on the assumption that there is a steady rate of increase of population throughout the year.

Red : This part is based on the assumption that there is a faster rate of increase of population in the second half-year.

For the first half-year the calculation is

number alive at time t_0	$= f(t_0)$
number born in the first half year	$\simeq \frac{1}{2}bf(t_0)$;
number dying in first half year	$\simeq \frac{1}{2}df(t_0)$;
number alive at time $t_0 + \frac{1}{2}$	$\simeq (1 + \frac{1}{2}b - \frac{1}{2}d)f(t_0)$,

so that

$$f(t_0 + \tfrac{1}{2}) \simeq (1 + \tfrac{1}{2}x)f(t_0)$$

By a similar calculation starting halfway through the year, we find:

$$f(t_0 + 1) \simeq (1 + \tfrac{1}{2}x)f(t_0 + \tfrac{1}{2});$$

substituting for $f(t_0 + \frac{1}{2})$ from our previous equation yields

$$f(t_0 + 1) \simeq (1 + \tfrac{1}{2}x)^2 f(t_0) \qquad \textbf{Equation (2)}$$

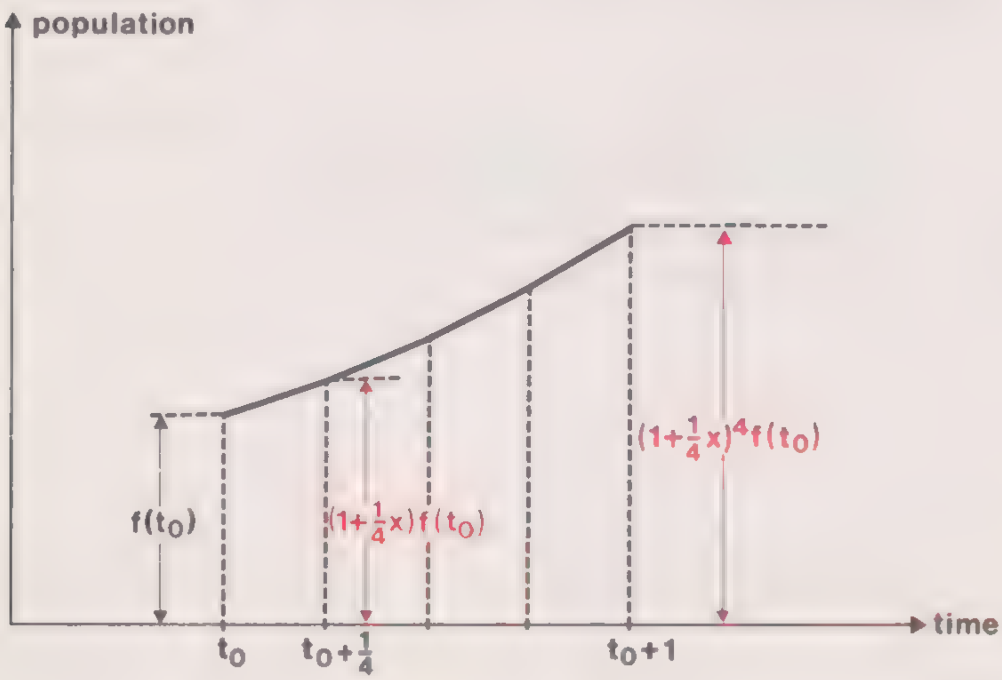
This is a more accurate result than Equation (1), though it is still approximate because we have assumed that there are as many births in the first quarter year as the second and as many in the third quarter as the fourth. To improve the approximation still further we could divide the year into four parts. Then a similar calculation to the one above gives

$$\left.\begin{aligned}
f(t_0 + \tfrac{1}{4}) &\simeq (1 + \tfrac{1}{4}x)f(t_0) \\
f(t_0 + \tfrac{1}{2}) &\simeq (1 + \tfrac{1}{4}x)f(t_0 + \tfrac{1}{4}) \\
f(t_0 + \tfrac{3}{4}) &\simeq (1 + \tfrac{1}{4}x)f(t_0 + \tfrac{1}{2}) \\
f(t_0 + 1) &\simeq (1 + \tfrac{1}{4}x)f(t_0 + \tfrac{3}{4})
\end{aligned}\right\}$$

Combining the four equations gives

$$f(t_0 + 1) \simeq (1 + \tfrac{1}{4}x)^4 f(t_0) \qquad \textbf{Equation (3)}$$

This is more accurate than Equation (2), though still approximate.



By further subdividing the year in this fashion into more and more parts we can obtain even better approximations to $f(t_0 + 1)$; subdividing the year into k equal parts gives

$$f(t_0 + 1) \simeq \left(1 + \frac{x}{k}\right)^k f(t_0)$$

Equation (4)

By making the number of subdivisions, k, very large, we would expect to get a very good approximation from this formula. The "intuitive" definition of a limit tells us that $f(t_0 + 1)$ is given as accurately as possible by the limit of the sequence of successive approximations.

In symbols, it tells us that

$$f(t_0 + 1) = \lim_{k \text{ large}} \left(1 + \frac{x}{k}\right)^k f(t_0)$$

This formula, which is exact within the restrictions of our model of population growth, solves the problem posed at the beginning of this section, by telling us that over 1 year the population increases by a factor which is the limit of the sequence

$$1 + x, (1 + \tfrac{1}{2}x)^2, (1 + \tfrac{1}{3}x)^3, \ldots$$

The importance of this limit goes far beyond the particular problem used here to introduce it. It has many applications in science, engineering and social science, as well as in mathematics itself. To give you an idea of how the sequence behaves, here are the first 10 elements in the cases $x = 0.1$ and $x = 1$.

k	$\left(1 + \frac{0.1}{k}\right)^k$	$\left(1 + \frac{1}{k}\right)^k$
1	1.1	2
2	1.1025	2.25
3	1.103370	2.370370
4	1.103813	2.441406
5	1.104081	2.488320
6	1.104260	2.521626
7	1.104389	2.546500
8	1.104486	2.565785
9	1.104561	2.581175
10	1.104622	2.593742

For $x = 0.1$, the sequence converges fairly rapidly and the limit is 1.105 to 3 decimal places. For $x = 1$ the convergence is slower, but by taking the calculation far enough we would obtain any desired accuracy. The value of the limit when $x = 1$ is a number whose frequency of occurrence in mathematical work rivals that of π. It is denoted by e, and to 5 decimal places its value is

$$e = 2.71828$$

The value of the limit for other values of x also appears very frequently. The function that maps x to the value of this limit is called the exponential function, and denoted by exp, so that

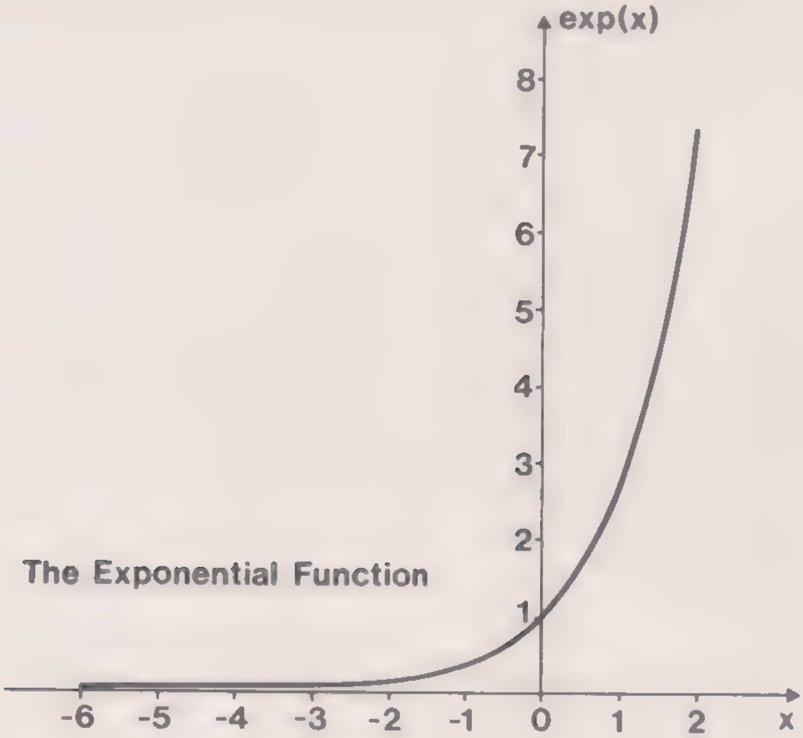
$$\exp : x \longmapsto \lim_{k \text{ large}} \left(1 + \frac{x}{k}\right)^k \qquad (x \in R)$$

and

$$\exp(x) = \lim_{k \text{ large}} \left(1 + \frac{x}{k}\right)^k \qquad (x \in R)$$

We often abbreviate $\exp(x)$ to $\exp x$.

The graph of this function is shown below.



The Exponential Function

We see from the graph that $\exp x$ is *positive* for *all* values of x.

Exercise 1

Use the information given on the last few pages to evaluate

$$\exp(0), \exp(1) \text{ and } \exp(\tfrac{1}{10})$$

to 3 decimal places.

5.2 The Exponential Theorem

You may have wondered why we used the name "exponential" for the function we discussed in the preceding section. The reason is that it is closely related to the idea of an exponent (an exponent is a number telling us to raise some other number to a power, for example the 2 in 5^2 or the 6 in 10^6). In this section we state a theorem which exhibits this relationship: we shall discuss the proof in section 5.4, together with other results stated in this chapter.

A special case of the theorem is

$$\exp\left(\frac{p}{q}\right) = e^{p/q} \quad (p \in Z, q \in Z^+)^*$$

Equation (1)

This means, for example, that

$$\exp(-1) = e^{-1} = \frac{1}{e}$$

that

$$\exp(\tfrac{1}{2}) = e^{1/2} = \sqrt{e}$$

and so on.

Exercise 1

Evaluate $\exp(2)$, working to two significant figures, using Equation (1) and the information contained in this text only.

The general statement of the exponential theorem is

$$\exp(x) = e^x \quad (x \in R)$$

Equation (2)

* For example, we write $-\dfrac{1}{2}$ as $\dfrac{-1}{2}$.

Even assuming that it is not difficult to prove the result expressed by Equation (1), there is an important new point in Equation (2): What do we mean by e^x when x is irrational? Since $\exp(x)$ is defined for all real x, we can use Equation (2) to give meaning to e^x. That is, for rational x we prove the result in Equation (2), and for irrational x we *define* e^x by Equation (2).

It is worth noticing that if x and y are rationals it follows from Equation (2) that

$$\exp(x + y) = e^{x+y} = e^x e^y = \exp(x)\exp(y)$$

by the laws of indices. In fact, the equation

$$\exp(x + y) = \exp(x)\exp(y)$$ Equation (3)

holds for *all* real numbers x and y

5.3 Natural Logarithms

We have seen how to define e^x for irrational x. How do we define a^x, where a is some real positive number other than e? One way to do it is to find a real number b such that

$$a = e^b = \exp(b)$$ Equation (1)

and then (remembering the laws of indices) to define

$$a^x = (e^b)^x = e^{bx} = \exp(bx)$$ Equation (2)

To evaluate a^x, therefore, we need the real number b; that is, we must solve Equation (1) by reversing the function exp. Just as the reverse of the function $x \longmapsto 10^x$ is a function called the logarithm to base 10, so the reverse of the function $\exp : x \longmapsto e^x$ is called the logarithm to base e or natural logarithm. The first tables of logarithms, made by John Napier in 1614, were a kind of natural logarithm. The symbol for the natural logarithm function is ln (or sometimes $\log_e$ or just log).

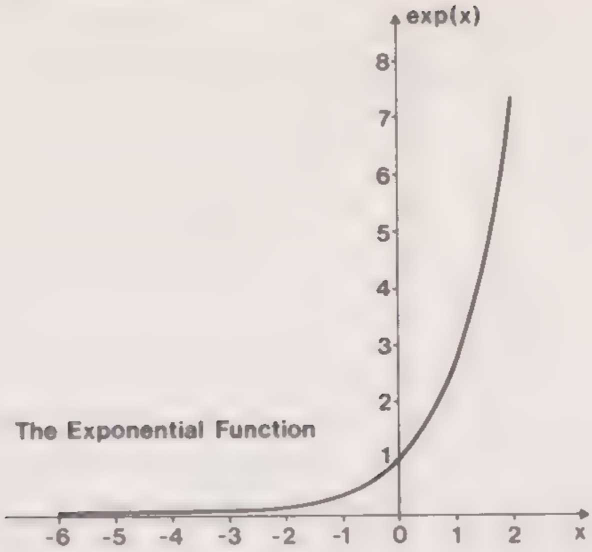
The graph of the exponential function, given again below, shows exp to be a one-one function with domain R and codomain R^+. Its reverse, the natural logarithm function, is therefore its inverse and a one-one mapping too, i.e. it is a function. Its graph is shown below. Notice how either graph can be obtained from the other by interchanging the x and y axes. This is a general characteristic of inverse and reverse mappings.

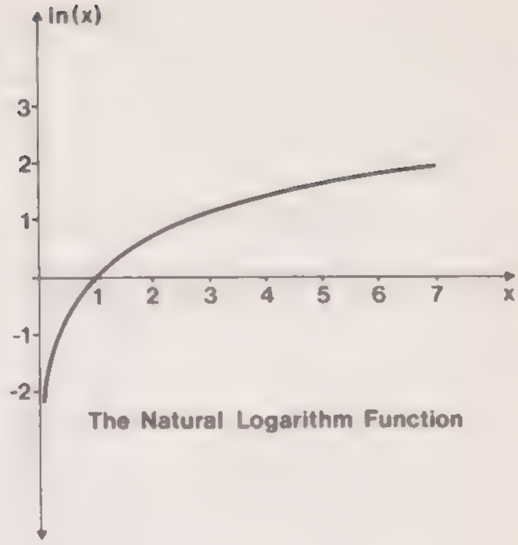
Exercise 1

(i) From the graphs given below, what are the domain and codomain of ln?

(ii) If $\ln(z) = \ln(x) + \ln(y)$, express z in terms of x and y.

(iii) Evaluate $\ln(e)$.





5.4 Two Formal Proofs

Proof that $\lim\limits_{k \text{ large}} \left(1 + \frac{x}{k}\right)^k$ exists

In section 5.1 we discussed intuitively the convergence of the sequence

$$1 + x, \left(1 + \frac{x}{2}\right)^2, \left(1 + \frac{x}{3}\right)^3, \ldots \quad (x \in R)$$

We made the implied assumption, however, that the limit of the sequence *exists*. We now give a formal proof of this, considering two cases according to the sign of x.

Case I: $x \geqslant 0$

We show that the elements of the sequence $k \longmapsto \left(1 + \frac{x}{k}\right)^k$ increase as k increases, and yet never exceed a fixed real number (such a fixed number is called an upper bound), so that the sequence must converge. By the binomial theorem we have

$$\begin{aligned}\left(1 + \frac{x}{k}\right)^k &= 1 + \frac{kx}{k} + \frac{k(k-1)}{k^2}\frac{x^2}{2!} + \frac{k(k-1)(k-2)}{k^3}\frac{x^3}{3!} + \cdots \\ &\quad + \frac{k(k-1)\ldots 1}{k^k}\frac{x^k}{k!} \\ &= 1 + x + \left(1 - \frac{1}{k}\right)\frac{x^2}{2!} + \left(1 - \frac{1}{k}\right)\left(1 - \frac{2}{k}\right)\frac{x^3}{3!} + \cdots \\ &\quad + \left(1 - \frac{1}{k}\right)\cdots\left(1 - \frac{k-1}{k}\right)\frac{x^k}{k!}\end{aligned}$$

If k is increased then the coefficient of each power of x increases, and in addition some new terms are added to the polynomial which are positive for positive x; so $\left(1+\frac{x}{k}\right)^k$ increases with k. To show that the elements of the sequence are bounded, let N be any integer greater than x; then for $k > N$ the above formula for $\left(1+\frac{x}{k}\right)^k$ gives

$$\left(1+\frac{x}{k}\right)^k \leqslant 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^N}{N!} + \frac{x^{N+1}}{(N+1)!} + \frac{x^{N+2}}{(N+2)!} + \cdots + \frac{x^k}{k!}$$

$$= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^{N-1}}{(N-1)!}\right) + \frac{x^N}{N!} \times \left(1 + \frac{x}{N+1} + \frac{x^2}{(N+1)(N+2)} + \ldots + \frac{x^{k-N}}{(N+1)\ldots k}\right)$$

$$\leqslant \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^{N-1}}{(N-1)!}\right) + \frac{x^N}{N!} \times \left(1 + \frac{x}{N} + \frac{x^2}{N^2} + \ldots + \frac{x^{k-N}}{N^{k-N}}\right)$$

$$= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^N}{N!}\left(\frac{1-\left(\frac{x}{N}\right)^{k-N+1}}{1-\left(\frac{x}{N}\right)}\right)$$

(after summing the geometric progression)

$$\leqslant 1 + x + \frac{x^2}{2!} + \frac{x^2}{3!} + \cdots + \frac{x^N}{N!}\left(\frac{1}{1-\left(\frac{x}{N}\right)}\right)$$

(since $N > x$ and $x \geqslant 0$)

which is independent of k, and is therefore an upper bound on every

element of the sequence. There is a theorem, which we shall not prove here, that any sequence whose elements increase with k but have an upper bound must converge. By this theorem, therefore, the sequence $\left(1 + \frac{x}{k}\right)^k$ converges for $x \geqslant 0$.

Case II: $x < 0$

We can prove the convergence of the sequence defining $\exp(x)$ as a by-product of the following result, which also serves to prove the multiplication theorem for the exponential function:

$$\exp(y - z) = \frac{\exp(y)}{\exp(z)} \qquad (y \text{ and } z \in R) \qquad \text{Equation (1)}$$

i.e. the sequence $k \longmapsto \left(1 + \frac{(y - z)}{k}\right)^k$ converges and its limit is $\frac{\exp(y)}{\exp(z)}$.

To show this we consider the expression

$$\frac{\left(1 + \frac{y - z}{k}\right)^k \left(1 + \frac{z}{k}\right)^k}{\left(1 + \frac{y}{k}\right)^k} = \left(\frac{1 + \frac{y}{k} + \frac{(y - z)z}{k^2}}{1 + \frac{y}{k}}\right)^k = \left(1 + \frac{\theta_k}{k}\right)^k$$

Equation (2)

where y and z are positive or zero, and

$$\theta_k = \frac{(y - z)z}{k\left(1 + \frac{y}{k}\right)} \qquad (k = 1, 2, \ldots)$$

The binomial expansion gives

$$\left(1 + \frac{\theta_k}{k}\right)^k - 1 = k\left(\frac{\theta_k}{k}\right) + \frac{k(k - 1)}{2}\left(\frac{\theta_k}{k}\right)^2 + \cdots + \left(\frac{\theta_k}{k}\right)^k,$$

and so

$$\left|\left(1 + \frac{\theta_k}{k}\right)^k - 1\right| \leqslant |\theta_k| + |\theta_k|^2 + \cdots + |\theta_k|^k$$

$$= |\theta_k|\left(\frac{1 - |\theta_k|^k}{1 - |\theta_k|}\right) \leqslant \frac{|\theta_k|}{1 - |\theta_k|}$$

provided that $|\theta_k| < 1$.

The definition of θ_k implies that $\lim\limits_{k \text{ large}} (\theta_k) = 0$ and hence that $\dfrac{|\theta_k|}{1 - |\theta_k|}$ can be made as small as we please by making k large enough. Consequently, $\left|\left(1 + \dfrac{\theta_k}{k}\right)^k - 1\right|$ can also be made as small as we please; and so the limit of the sequence $\left(1 + \dfrac{\theta_k}{k}\right)^k$ is 1. It follows, by the multiplication rule for limits, see section 6.2, and by Equation (2), that

$$\lim_{k \text{ large}} \left(1 + \frac{y - z}{k}\right)^k = \frac{\lim\limits_{k \text{ large}} \left(1 + \dfrac{y}{k}\right)^k \lim\limits_{k \text{ large}} \left(1 + \dfrac{\theta_k}{k}\right)^k}{\lim\limits_{k \text{ large}} \left(1 + \dfrac{z}{k}\right)^k}$$

(Note that the denominator is non-zero.)

so

$$\exp(y - z) = \frac{\exp(y)}{\exp(z)}$$

We can use this result in two ways. First, taking $y = 0$ and z positive, we have a proof that the sequence defining $\exp(z)$ for negative values of z converges, and that its limit is the reciprocal of $\exp(-z)$. That is,

$$\exp(-z) = \frac{1}{\exp(z)}$$

Secondly, we can use it to prove the result:

$$\exp(x_1)\exp(x_2) = \exp(x_1 + x_2)$$

by substituting, e.g., $y = x_1 + x_2$, $z = x_1$ in Equation (1), where $x_1 \geqslant 0$ and $x_2 \geqslant 0$.

We will now recapitulate the main theorem. We had previously defined the function exp (if it exists) by

$$\exp(x) = \lim_{k \text{ large}} \left(1 + \frac{x}{k}\right)^k \quad (x \in R)$$

We have just proved that this limit exists, and therefore the function exists with domain R. In particular we may also define a constant

$$e = \exp(1) = \lim_{k \text{ large}} \left(1 + \frac{1}{k}\right)^k$$

We have incidentally provided a sound basis for the next theorem.

Proof of the Exponential Theorem

We have discussed the *Exponential Theorem* in section 5.2. The general statement of the theorem is

$$\exp(x) = e^x \quad (x \in R)$$

and we now give a proof of this. It starts with a lemma about $\exp(x/k)$ where x is a real number and k a positive integer. We know that the sequence defining $\exp(x)$ is:

$$1 + x, \quad (1 + \tfrac{1}{2}x)^2, \quad (1 + \tfrac{1}{3}x)^3, \quad (1 + \tfrac{1}{4}x)^4, \ldots$$

For $k = 2$, the sequence defining $\exp\left(\dfrac{x}{k}\right)$ is

$$1 + \tfrac{1}{2}x, \quad (1 + \tfrac{1}{4}x)^2, \quad (1 + \tfrac{1}{6}x)^3, \ldots$$

and so by the multiplication rule for limits, see Section 6.2,

$\left(\exp\left(\dfrac{x}{2}\right)\right)^2$ is the limit of the sequence

$$(1 + \tfrac{1}{2}x)^2, \quad (1 + \tfrac{1}{4}x)^4, \ldots$$

This sequence consists of alternate elements from the first sequence above; and since the terms $\left(1 + \dfrac{x}{n}\right)^n$ for n large in the first sequence are all close to its limit, the ones which appear in the last sequence are close to this limit, and so the last sequence has the same limit as the first. This demonstrates that

$$\left(\exp\left(\frac{x}{2}\right)\right)^2 = \exp(x)$$

For a general positive integer k, the demonstration is analogous; in place of the last sequence we get

$$\left(1 + \frac{1}{k}x\right)^k, \quad \left(1 + \frac{1}{2k}x\right)^{2k}, \ldots$$

which consists of every kth term from the first sequence and therefore has the same limit.

The lemma just demonstrated,

$$\exp x = \left(\exp\left(\frac{x}{k}\right)\right)^k \quad (x \in R \text{ and } k \in Z^+)$$

can be used in two ways. First we set $x = k = p$ and obtain

$$\exp(p) = (\exp(1))^p = e^p \quad (p \in Z^+)$$

Secondly, we set $x = p$ and $k = q$ in the same result, with p and q both positive integers, obtaining

$$\exp(p) = \left(\exp\left(\frac{p}{q}\right)\right)^q \quad (p, q \in Z^+)$$

Substituting for $\exp(p)$ and then interchanging the two sides of the equation:

$$\left(\exp\left(\frac{p}{q}\right)\right)^q = e^p \quad (p, q \in Z^+)$$

Finally, taking the qth root of both sides, we obtain

$$\exp\left(\frac{p}{q}\right) = \sqrt[q]{e^p}$$

$$= e^{p/q} \text{ by the laws of indices.}$$

This is equivalent to the statement that

$$\exp(x) = e^x \qquad \text{Equation (3)}$$

when x is positive and rational. This result cries out to be generalized to negative values and irrational values of x. The generalization to negative values depends on the theorem

$$\exp(x) \times \exp(y) = \exp(x + y) \quad (x \in R \text{ and } y \in R) \qquad \text{Equation (4)}$$

whose proof has been indicated earlier. If x is negative and rational, we put $y = -x$ and so Equations (3) and (4) give

$$\exp(x) = \frac{\exp(0)}{\exp(y)}$$

$$= \frac{1}{e^y} \qquad \text{by Equation (3)} \quad (\text{since } y > 0)$$

$$= e^{-y} = e^x$$

This proves the exponential theorem for negative rational x.

The generalization of the exponential theorem to irrational values of x depends on how we define e^x. If x is rational, it can be put in the form $\frac{p}{q}$ with p and q integers, and then e^x means $e^{p/q}$, that is the qth root of e^p; but we have no corresponding definition for irrational values of x. The problem for irrational x, therefore, is not how to prove Equation 5.2. (2) but how to define e^x; the natural answer is to adopt Equation 5.2 (2)

as our definition of e^x for irrational x. Combining our previous results with this definition for irrational x, we end up with the important result

$$\exp(x) = e^x \quad (x \in R)$$

which is the *exponential theorem.*

5.5 Additional Exercises

Exercise 1

A savings bank offers the rate of interest $r\%$ compounded annually. A man deposits some money and leaves it for several years. If the amount of money to his credit after k years is £u_k, write down a recurrence formula for the sequence $u_1, u_2, \ldots$. (Assume that this bank works with exact arithmetic instead of approximating by whole-penny amounts as real banks do.)

What is the recurrence formula relating u_k to u_{k-1} if interest at the rate of $r\%$ per annum is compounded

(i) half-yearly,
(ii) quarterly,
(iii) monthly?

Exercise 2

Investigate $\exp \circ \ln$ and $\ln \circ \exp$. Do we get the same function in each case?

5.6 Answers to Exercises

Section 5.1

Exercise 1

The definition of $\exp(x)$ gives

$$\exp(0) = \lim_{k \text{ large}} \left(1 + \frac{0}{k}\right)^k = \lim(1) = 1$$

$$\exp(1) = \lim_{k \text{ large}} \left(1 + \frac{1}{k}\right)^k = e = 2.718$$

$$\exp\left(\frac{1}{10}\right) = \lim_{k \text{ large}} \left(1 + \frac{0.1}{k}\right)^k = 1.105$$

(This last result can be deduced from the table on page 103.)

The first two of these results, $\exp(0) = 1$ and $\exp(1) = e$, are important.

Section 5.2

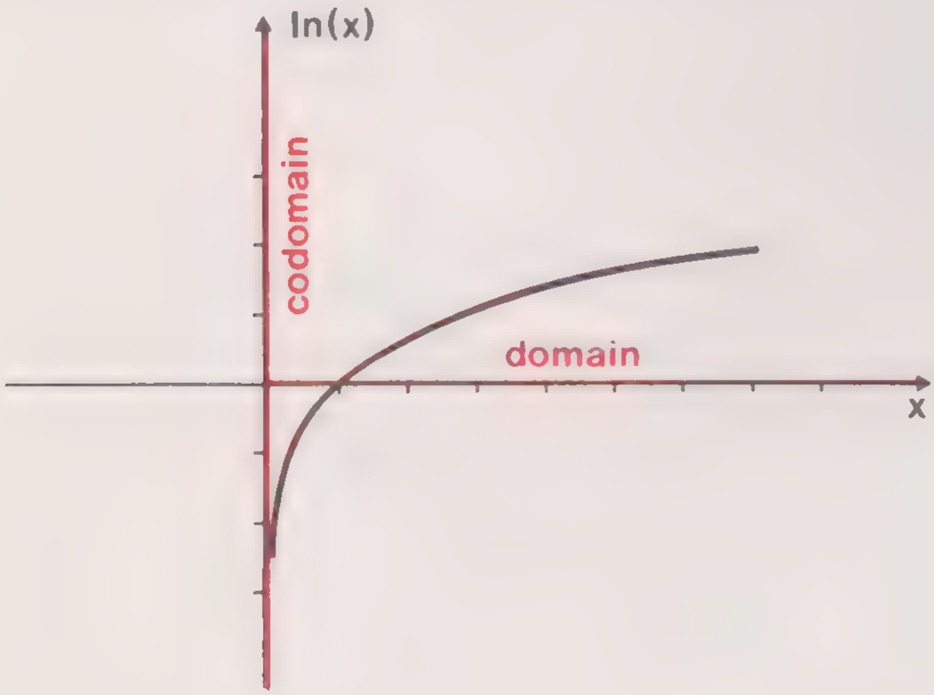
Exercise 1

$$\begin{aligned}\exp(2) &= e^2\\ &= (2.72)^2\\ &= 7.4 \text{ to the required accuracy}\end{aligned}$$

(The true value is 7.39 to two decimal places.)

Section 5.3

Exercise 1



(i) Domain R^+, codomain R.

This can be seen from the graph. Since ln is the inverse of exp, its domain is the image set of exp: $\exp(x)$ gets very small when x is large and negative, but it is never actually zero. The domain of ln is therefore R^+: only positive numbers have logarithms. The codomain of ln is the domain of exp, and is therefore R.

(ii) $z = xy$

To find z, given $\ln(z)$, we apply the inverse mapping exp which gives

$$\begin{aligned} z &= \exp(\ln(z)) && \text{by definition of ln} \\ &= \exp(\ln(x) + \ln(y)) && \text{by given data} \\ &= \exp(\ln(x)) \times \exp(\ln(y)) && \text{by Equation 5.2.(3).} \\ &= xy && \text{by definition of ln} \end{aligned}$$

(iii) $$\begin{aligned} \ln(e) &= \ln(\exp(1)) && \text{by definition of } e \\ &= 1 && \text{by definition of ln} \end{aligned}$$

Section 5.5

Exercise 1

At the beginning of the kth year the account contains £u_{k-1}. The amount of interest credited during the year is $\frac{r}{100}u_{k-1}$, and so the balance at the end of the kth year is $u_{k-1} + \frac{r}{100}u_{k-1}$. Accordingly the recurrence relation is

$$u_k = \left(1 + \frac{r}{100}\right)u_{k-1}$$

For interest compounded half-yearly, monthly and quarterly (respectively), we have:

(i) $u_k = \left(1 + \frac{r}{200}\right)^2 u_{k-1}$

(ii) $u_k = \left(1 + \frac{r}{400}\right)^4 u_{k-1}$

(iii) $u_k = \left(1 + \frac{r}{1200}\right)^{12} u_{k-1}$

Exercise 2

If f and g are functions:

$$f: x \longmapsto \exp(x) \quad (x \in R),$$
$$g: x \longmapsto \ln(x) \quad (x \in R^+),$$

then since the codomain of g is the same as the domain of f, $f \circ g$ is defined and is the identity mapping

$$x \longmapsto x \quad (x \in R^+).$$

Similarly, since the codomain of f is the same as the domain of g, $g \circ f$ is defined and is the identity mapping

$$x \longmapsto x \quad (x \in R).$$

Thus $f \circ g \neq g \circ f$ since the domain of $f \circ g$ is not the same as that of $g \circ f$.

CHAPTER 6 CONVERGENCE

6.0 Introduction

In this chapter we develop the concept of *limit* further, and we present precise definitions which can replace the intuitive definitions of Chapters 2 and 4. We begin by again looking at sequences and after giving a rigorous definition of the limit of a sequence, we introduce some formal concepts.

We make use of the idea of the limit of a sequence to introduce limits of series, and we discuss the convergence of infinite series.

Finally, we give rigorous definitions of the limit of a real function and of *continuity*.

6.1 Limits of Sequences

In Chapter 2 we denoted the infinite sequence $u_1, u_2, u_3, \ldots$ by $\underset{\sim}{u}$ and its kth term by u_k. We then gave the following definition:

> **Intuitive Definition of a Limit**
>
> "The number $\lim \underset{\sim}{u}$ is the limit of the infinite sequence $\underset{\sim}{u}$" is equivalent to the statement "if k is very large, then u_k is a very good approximation to $\lim \underset{\sim}{u}$".

This definition can, however, lead to ambiguities if it is pushed too far.

One reason why the intuitive definition of a limit leads to ambiguities is that the phrases "k is very large" and "u_k is a very good approximation to $\lim \underset{\sim}{u}$" have not been defined. In fact, these are phrases whose meanings depend on the circumstances: what seems large to a mouse may not seem large to an elephant; to a butcher weighing meat 0.499 kg may seem a good approximation to 0.5 kg but a pharmacist measuring out a dangerous drug will want to account for every 0.001 kg.

If u_k is *very close* to $\lim \underset{\sim}{u}$, then it is also very close to any other number that is also close to $\lim \underset{\sim}{u}$. In other words, it is very close to a *range of values*. Unless a more precise definition is provided we are not really in a position to talk about *the* limit of a sequence.

We can interpret the statement "u_k is a very good approximation to $\lim \underset{\sim}{u}$" to mean that the difference between u_k and $\lim \underset{\sim}{u}$ is less than or equal to a positive number, ε. In symbols, this can be written

$$|u_k - \lim \underset{\sim}{u}| \leqslant \varepsilon$$

or equivalently

$$-\varepsilon \leqslant u_k - \lim \underset{\sim}{u} \leqslant \varepsilon$$

Another way of writing the same condition, obtained by adding $\lim \underset{\sim}{u}$ to all members of the inequalities, is

$$\lim \underset{\sim}{u} - \varepsilon \leqslant u_k \leqslant \lim \underset{\sim}{u} + \varepsilon$$

which is the same as

$$u_k \in [\lim \underset{\sim}{u} - \varepsilon, \lim \underset{\sim}{u} + \varepsilon]$$

where

$$[\lim \underset{\sim}{u} - \varepsilon, \lim \underset{\sim}{u} + \varepsilon]$$

stands for the set of all real numbers x satisfying

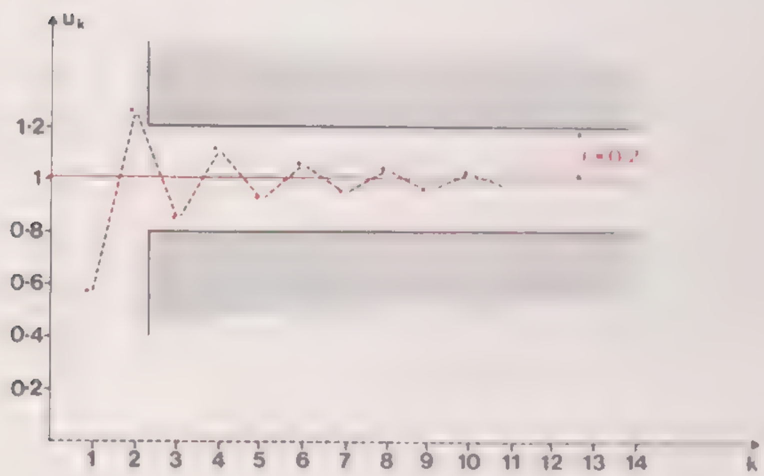
$$\lim \underset{\sim}{u} - \varepsilon \leqslant x \leqslant \lim \underset{\sim}{u} + \varepsilon$$

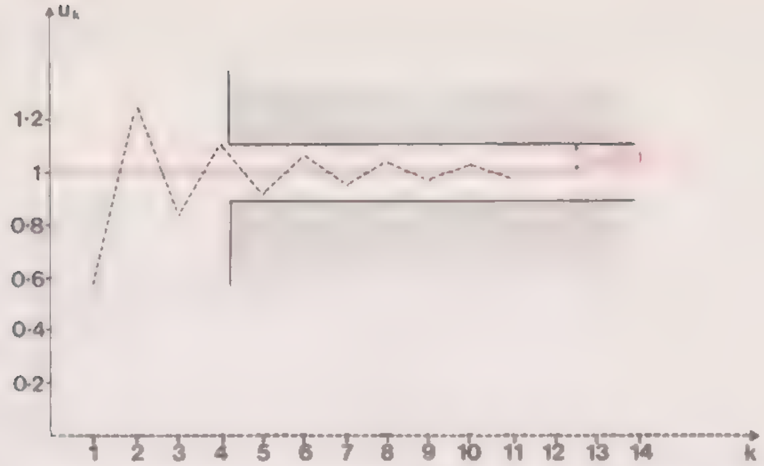
We call this set the error interval associated with the limit $\lim \underset{\sim}{u}$ and the error bound ε. (Thus if the limit is 1 and the error bound is 0.005, then the error interval comprises all real numbers from 0.995 to 1.005 inclusive.)

We now want to frame a definition for "k is very large" that gives a similar precision to this statement. To see how this can be done, look at the following diagrams, where the graph of the sequence

$$u_k = 1 + \frac{(-1)^k 4}{(k+2)^2} \quad (k \in Z^+)$$

is plotted.





In the first diagram the only points of the graph that do not lie between the parallel lines representing the error interval $[1 - 0.2, 1 + 0.2]$ are those with $k = 1$ and $k = 2$; consequently, with $\varepsilon = 0.2$, we can say that the "very large" values of k start at 3. In the second diagram the only points which are not between the parallel lines representing the error interval $[1 - 0.1, 1 + 0.1]$ are the ones with $k = 1, 2, 3$ and 4; consequently, with $\varepsilon = 0.1$, the "very large" values of k start at 5. If ε is made smaller, more and more k values get left outside the parallel lines, but it is always possible to pick a point on the graph (like the point $(3, u_3)$ for $\varepsilon = 0.2$ or $(5, u_5)$ for $\varepsilon = 0.1$) to the right of which *all* the points of the graph lie between the parallel lines. This is the essential part: for *any* positive value of the error bound ε, there is an element of the sequence after which all the elements lie in the error interval

$$[1 - \varepsilon, 1 + \varepsilon]$$

For example, if ε is $\dfrac{1}{10^6}$, then the condition for an element u_k to lie in the error interval is

$$1 - \frac{1}{10^6} \leqslant 1 + \frac{(-1)^k 4}{(2 + k)^2} \leqslant 1 + \frac{1}{10^6}$$

We can subtract 1 from every member of these inequalities, obtaining

$$-\frac{1}{10^6} \leqslant \frac{(-1)^k 4}{(2 + k)^2} \leqslant \frac{1}{10^6}$$

Since $(-1)^k$ take the values $+1$ and -1 only, this pair of inequalities is satisfied if

$$-\frac{1}{10^6} \leqslant \frac{-4}{(2 + k)^2} \quad \text{and} \quad \frac{4}{(2 + k)^2} \leqslant \frac{1}{10^6}$$

Since these last two inequalities are equivalent, it is sufficient to require

$$\frac{4}{(2+k)^2} \leqslant \frac{1}{10^6}$$

which is equivalent to $4\,000\,000 \leqslant (2+k)^2$, and is therefore true for all $k \geqslant 1998$. If ε is $\frac{1}{10^6}$, then all elements after the 1998th lie in the error interval.

Generalizing this idea to any infinite sequence, we can now adopt the following definition:

> **Rigorous Definition of a Limit**
>
> "The number $\lim \underset{\sim}{u}$ is the limit of the infinite sequence $\underset{\sim}{u}$" is equivalent to the statement "for any positive error bound ε, there is an element of the sequence after which every element lies within the error interval $[\lim \underset{\sim}{u} - \varepsilon, \lim \underset{\sim}{u} + \varepsilon]$."

In effect, this definition states that, whatever accuracy we choose to work with, we can always use $u_1, u_2, u_3, \ldots$ as a sequence of successive approximations for calculating the number $\lim \underset{\sim}{u}$ (if $\lim \underset{\sim}{u}$ exists); for there is an element in the sequence beyond which all the elements are (to this accuracy) indistinguishable from each other, and so any of these elements may be used as the calculated approximate value of $\lim \underset{\sim}{u}$.

It takes quite a lot of experience to get used to the definition. If you are unhappy about it, look again at the two diagrams and try to imagine how they would look if ε were still further reduced, and whether a suitable value of N (the distance from the u_k-axis to the shaded area) could be found however small ε were chosen. In doing this, remember that the graph we have drawn only shows the first few elements in the sequence, but that the true graph extends indefinitely to the right, since the sequence has no last element. (Indeed, the concepts of a limit and of convergence do not apply to finite sequences.)

One of the difficulties in using the rigorous definition of a limit is that it requires us to prove something about *any* positive value of ε, and about *every* element after the Nth in the sequence. We cannot deal individually with each value of ε or every element of the sequence; so instead we must find a way of dealing with them all at once. In effect we have to prove a tiny theorem for each individual sequence. The following example shows how it is done.

Example 1

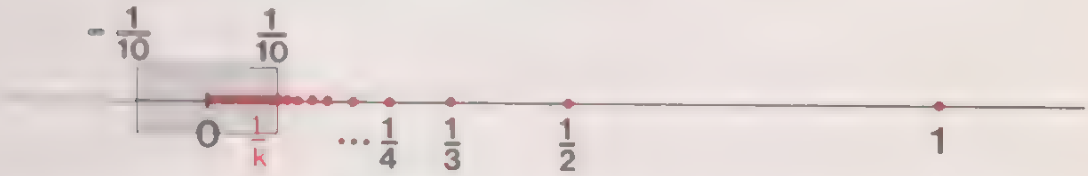
Show that 0 is the limit of the sequence $1, \frac{1}{2}, \frac{1}{3}, \ldots$.

The definition requires us to show that for any positive error bound ε there is an element (say the Nth) after which all elements are within the error interval $[0 - \varepsilon, 0 + \varepsilon]$. As a start, let us show it for a particular value of ε, say $\frac{1}{10}$. Then, since the kth element of the sequence is $\frac{1}{k}$, it is a question of finding a positive integer N such that all elements u_k after the Nth satisfy

$$-\frac{1}{10} \leqslant u_k \leqslant \frac{1}{10}$$

That is, we wish to find $N \in Z^+$ such that

$$-\frac{1}{10} \leqslant \frac{1}{k} \leqslant \frac{1}{10} \quad (k \in Z^+ \text{ and } k > N)$$



Now the inequality $-\frac{1}{10} \leqslant \frac{1}{k}$ is satisfied for all $k \in Z^+$, so that it implies no restriction on N. Multiplying both sides of the inequality by $10k$, (a step that is justified since $k \in Z^+$), we find that it is equivalent to

$$10 \leqslant k \quad (k \in Z^+ \text{ and } k > N)$$

That is, we want to find a positive integer N such that every integer k greater than N is also greater than or equal to 10. There are many numbers N with this property; one of them is 10 itself. So we have shown that if ε has the particular value $\frac{1}{10}$ we can satisfy the definition of a limit for this sequence by taking $N = 10$.

To complete the proof that 0 is the limit of the sequence $u_k = \frac{1}{k}$ we must show that for *any* positive ε, not necessarily $\frac{1}{10}$, we can find an N that satisfies the definition. That is, we must find a positive integer N such that

$$-\varepsilon \leqslant \frac{1}{k} \leqslant \varepsilon \quad (k \in Z^+ \text{ and } k > N)$$

Once again the inequality $-\varepsilon \leqslant \frac{1}{k}$ places no restriction on N. The inequality $\frac{1}{k} \leqslant \varepsilon$ is equivalent to

$$k \geqslant \frac{1}{\varepsilon} \qquad (k \in Z^+ \text{ and } k > N)$$

and so we want an integer N such that every integer k greater than N is also greater than $\frac{1}{\varepsilon}$ (which must be positive, but need not be an integer). Such an N can indeed be found; for example, the first integer after $\frac{1}{\varepsilon}$ will do. Since we can find a suitable N for *any positive ε, however small*, the definition of a limit is satisfied, and so 0 is proved to be the limit of the sequence $u_k = \frac{1}{k}$.

Exercise 1

Verify that the limit of the sequence defined by

$$u_k = \frac{1}{k^2}$$

is 0.

You can take N to be the first integer after the number $\frac{1}{\sqrt{\varepsilon}}$.

Exercise 2

Verify that the limit of the sequence 0.3, 0.33, 0.333, 0.3333, . . . is $\frac{1}{3}$. You can use the fact that

$$\tfrac{1}{3} - 0.\underbrace{33\ldots3}_{k \text{ digits}} = 0.\underbrace{00\ldots0}_{k \text{ zeros}}333\ldots = \tfrac{1}{3} \times 10^{-k}$$

and, for any particular ε, take N to be the number of consecutive zeros after the decimal point in the decimal representation of ε.

6.2 Addition and Multiplication of Sequences

The object of this section is to develop a technique for simplifying the calculation of limits, so that it becomes unnecessary to go right back to the definition of a limit every time we want to evaluate one. The basic idea is that sequences with a complicated rule of formation can be built out of simpler ones by means of operations such as addition and multiplication. If we know how these algebraic operations are reflected in the behaviour of the limits of the sequences, we can evaluate the limits of more complicated sequences in terms of those of simpler ones. In other words we wish to define such things as the sum and the product of two sequences, and (for example) relate the limit of the sum of two sequences to the limits of the individual sequences.

The first algebraic operation we consider is addition. To arrive at a sensible definition for addition of sequences, we use the fact that an infinite sequence is specified by a function with domain Z^+ : that is to say, $\underline{u}$ is specified by the function f where

$$f : k \longmapsto u_k \qquad (k \in Z^+)$$

and $\underline{v}$ is specified by the function g where

$$g : k \longmapsto v_k \qquad (k \in Z^+)$$

The point of this observation is that we have already given, in Chapter 3, the definition for the addition of two functions. Applied to the two functions f and g with domains Z^+, the definition is

$$f + g : k \longmapsto f(k) + g(k) \quad (k \in Z^+)$$

We denote the sequence specified by $f + g$ by $\underline{u} + \underline{v}$; so $\underline{u} + \underline{v}$ is the sequence

$$u_1 + v_1, u_2 + v_2, u_3 + v_3 \ldots$$

That is, to add sequences, we add corresponding elements. For example, if $\underline{u}$ is $1, 2, 3, 4, \ldots$ and $\underline{v}$ is $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$, then $\underline{u} + \underline{v}$ is $2, 2\frac{1}{2}, 3\frac{1}{3}, 4\frac{1}{4}, \ldots$.

Note that the "+" in $\underline{u} + \underline{v}$ is really a *new* symbol, denoting the operation "addition of sequences". We use the same symbol as that for the operation "addition of real numbers" because the two operations have similar properties.

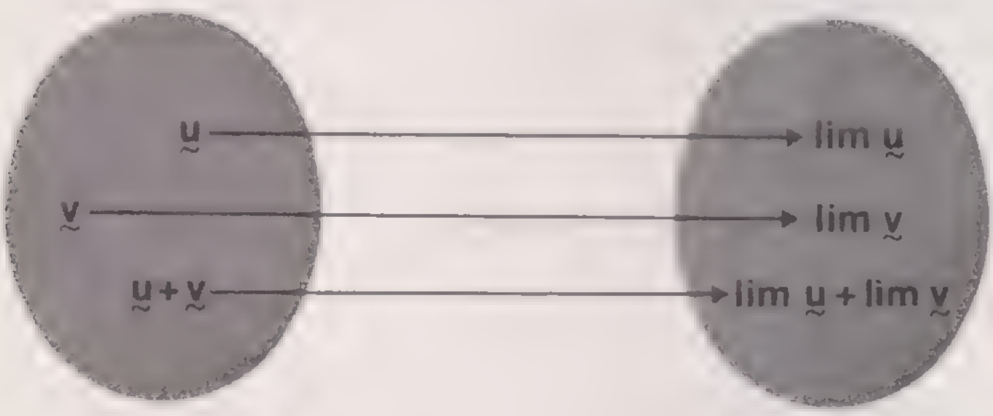
Exercise 1

If $\underset{\sim}{u}$ is $3\frac{1}{2}, 3\frac{1}{3}, 3\frac{1}{4}, 3\frac{1}{5}, \ldots$ and $\underset{\sim}{v}$ is $2, 2, 2, 2, \ldots$ what is $\underset{\sim}{u} + \underset{\sim}{v}$? Also, what are the limits of $\underset{\sim}{u}$, $\underset{\sim}{v}$, and $\underset{\sim}{u} + \underset{\sim}{v}$, and how are they related?

Exercise 2

If $\underset{\sim}{u}$ and $\underset{\sim}{v}$ are convergent sequences, give a demonstration (i.e. an argument based on the intuitive definition of a limit) that $\underset{\sim}{u} + \underset{\sim}{v}$ is also convergent and that its limit is $\lim \underset{\sim}{u} + \lim \underset{\sim}{v}$.

The result of the last exercise shows two things. First, it shows that the sum of two convergent sequences is another convergent sequence. Secondly, it shows that the operation of adding sequences is carried over by the function lim into the operation of adding numbers (the limits of these sequences). The same fact can be stated schematically as follows:



or by $\lim (\underset{\sim}{u} + \underset{\sim}{v}) = \lim \underset{\sim}{u} + \lim \underset{\sim}{v}$

We can prove this result rigorously as follows:

Proof that, if $\underset{\sim}{u}$ and $\underset{\sim}{v}$ converge, then $\underset{\sim}{u} + \underset{\sim}{v}$ converges and has the limit $\lim \underset{\sim}{u} + \lim \underset{\sim}{v}$.

(i) "$\underset{\sim}{u}$ is convergent" means that we can find $N \in Z^+$ such that u_k is as close to $\lim \underset{\sim}{u}$ as we please for all $k > N$.

(ii) "$\underset{\sim}{v}$ is convergent" means that we can find $M \in Z^+$ such that v_k is as close to $\lim \underset{\sim}{v}$ as we please for all $k > M$.

We wish to show that we can find $P \in Z^+$ such that $u_k + v_k$ is an approximation to $\lim \underset{\sim}{u} + \lim \underset{\sim}{v}$ with error bound less than or equal to ε, for any small positive number ε, and for all $k > P$:

We are *adding* the approximation u_k to $\lim \underset{\sim}{u}$ to the approximation v_k to $\lim \underset{\sim}{v}$ to obtain the approximation $u_k + v_k$ to $\lim \underset{\sim}{u} + \lim \underset{\sim}{v}$.

The error bound for a sum of approximations is equal to the sum of the error bounds of the individual approximations.

We therefore require that the sum of the error bounds of the two approximations be $\leqslant \varepsilon$. There is no reason for one of the error bounds to be less than the other, so we take them to be equal.

By (i), we can find $N \in Z^+$ such that the approximation u_k to $\lim \underset{\sim}{u}$ has error bound $\leqslant \frac{\varepsilon}{2}$ for all $k > N$: that is

$$u_k \in [\lim \underset{\sim}{u} - \tfrac{1}{2}\varepsilon, \lim \underset{\sim}{u} + \tfrac{1}{2}\varepsilon] \quad \text{when } k > N$$

By (ii), we can find $M \in Z^+$ such that the approximation v_k to $\lim \underset{\sim}{v}$ has error bound $\leqslant \frac{\varepsilon}{2}$ for all $k > M$; that is

$$v_k \in [\lim \underset{\sim}{v} - \tfrac{1}{2}\varepsilon, \lim \underset{\sim}{v} + \tfrac{1}{2}\varepsilon] \quad \text{when } k > M$$

Let P be the larger of M and N. The last two statements together imply that

$$u_k + v_k \in [\lim \underset{\sim}{u} + \lim \underset{\sim}{v} - \varepsilon, \lim \underset{\sim}{u} + \lim \underset{\sim}{v} + \varepsilon] \quad \text{for all } k > P$$

which is what we wished to prove, as this statement is true for *any* choice of ε.

The operation of multiplication can be dealt with in the same way. To define the multiplication of sequences, we again refer back to Chapter 3, where the product of two functions f and g with domain Z^+ is defined by:

$$f \times g : k \longmapsto f(k) \times g(k) \quad (k \in Z^+)$$

Denoting by $\underset{\sim}{u} \times \underset{\sim}{v}$ the sequence specified by $f \times g$, we see that $\underset{\sim}{u} \times \underset{\sim}{v}$ is the sequence

$$u_1 \times v_1, u_2 \times v_2, u_3 \times v_3, \ldots$$

i.e. to "multiply" sequences, we multiply corresponding elements. For example, if $\underset{\sim}{u}$ is 1, 2, 3, 4, ... and $\underset{\sim}{v}$ is 10, 100, 1000, 10 000, ..., then $\underset{\sim}{u} \times \underset{\sim}{v}$ is 10, 200, 3000, 40 000, ...

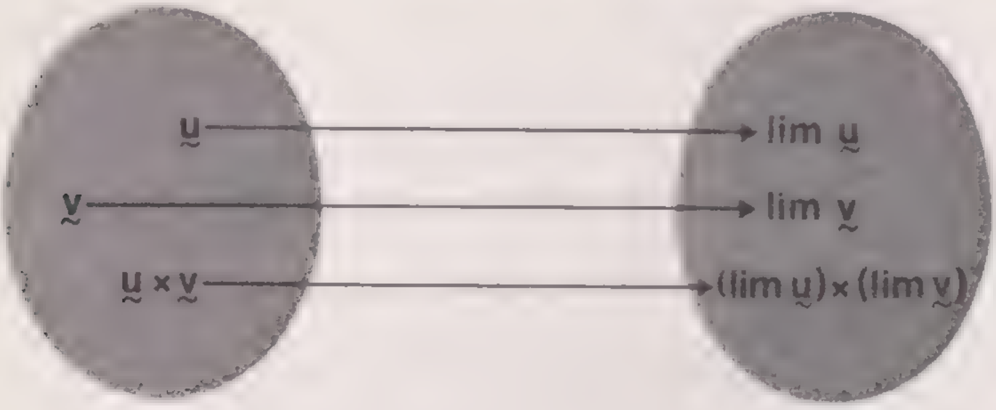
Note that the "$\times$" in $\underset{\sim}{u} \times \underset{\sim}{v}$ is really a *new* symbol, denoting the operation "multiplication of sequences". We use the same symbol as that for the operation "multiplication of real numbers" because the two operations have similar properties.

Exercise 3

If $\underset{\sim}{u}$ is 0.2, 0.22, 0.222, 0.2222, ... and $\underset{\sim}{v}$ is 3.3, 3.03, 3.003, 3.0003, ... what is $\underset{\sim}{u} \times \underset{\sim}{v}$? Also, what are the limits of $\underset{\sim}{u}$, $\underset{\sim}{v}$, and $\underset{\sim}{u} \times \underset{\sim}{v}$, and how are they related?

Exercise 4

If $\underset{\sim}{u}$ and $\underset{\sim}{v}$ are convergent sequences, give a demonstration that $\underset{\sim}{u} \times \underset{\sim}{v}$ is also convergent and that its limit is $(\lim \underset{\sim}{u}) \times (\lim \underset{\sim}{v})$.



6.3 Infinite Series

We have been considering *sequences* by looking at the terms individually. Now let us try adding them together. What we obtain is called a *series*.

A series containing a finite number of terms is called a *finite series*. Thus

$$a_1 + a_2 + a_3 + \cdots + a_{10},$$

is a finite series, and we can always add the terms of such a series to find its sum. If the number of terms is not finite, then we have an *infinite series*, and we need to consider what interpretation we now give to the sum of an infinite series.

It is very important to understand just what we mean by an infinite series. We give these important definitions formally:

An **infinite series** is an expression of the form

$$a_1 + a_2 + a_3 + \cdots.$$

The **partial sums** of the infinite series are the sums:

$$S_k = a_1 + a_2 + \cdots + a_k \quad (k = 1, 2, 3, \ldots).$$

If the sequence of partial sums,

$$S_1, S_2, S_3, \ldots$$

converges to a limit S, then we say that the series **converges (or is convergent) to the sum S**, and we write

$$S = a_1 + a_2 + a_3 + \cdots.$$

If the sequence of partial sums does not converge, then we say that the series diverges (or is divergent): we cannot find a sum for it.

It is important to note the difference between the infinite series

$$a_1 + a_2 + a_3 + \cdots$$

and the infinite sequence

$$a_1, a_2, \ldots$$

Example 1

You may have met the formula for the sum of k terms of a geometric progression,

$$a + ar + ar^2 + \cdots + ar^{k-1} = a\left(\frac{1 - r^k}{1 - r}\right) \qquad (r \in R, r \neq 1).$$

This is the kth partial sum, S_k, of the infinite series

$$a + ar + ar^2 + \cdots.$$

This series is called the infinite geometric series; the number r is called the common ratio

As an example, let us take $a = 1$ and $r = \frac{1}{2}$: then we have:

$$\begin{aligned} 1 + \tfrac{1}{2} + \tfrac{1}{4} + \cdots + 2^{-(k-1)} &= \frac{1 - 2^{-k}}{\frac{1}{2}} \\ &= 2 - 2^{-(k-1)}. \end{aligned}$$

Thus the sequence $S_1, S_2, S_3, \ldots$ is now

$$2 - 1, 2 - \tfrac{1}{2}, 2 - \tfrac{1}{4}, \ldots$$

which converges to 2, so we can write

$$1 + \tfrac{1}{2} + \tfrac{1}{4} + \cdots = 2.$$

Exercise 1

Obtain a formula for the kth partial sum of

$$1 - 1 + 1 - 1 + 1 \ldots$$

Does the series converge or diverge?

Exercise 2

For what values of r can we define a sum for the infinite geometric series

$$1 + r + r^2 + r^3 + \cdots$$

and what is the formula for the sum in each case?

6.4 Limits of Functions

We introduced two intuitive definitions of the *limit of a real function* in Chapter 4. Unlike a sequence, the image set of a real function of x may approach a limit as x approaches *any* number, and not just when x is very large. Care will be needed to specify which kind of limit is under discussion, e.g.

$$\lim_{x \text{ large}} \quad \text{or} \quad \lim_{x \to a}$$

In Chapter 4 we gave intuitive definition for the limit of a function.

For $\lim\limits_{x \text{ large}}$ we had:

Intuitive Definition of a Limit

If f is a real function and L is a number, then "L is the limit of f for large numbers in its domain" is equivalent to the statement "whenever x is very large, $f(x)$ is a very good approximation to L".

For $\lim\limits_{x \to a}$ we had:

Intuitive Definition of a Limit

If g is a real function and a and L are real numbers, "L is the limit of g near a" is equivalent to the statement "if x is very close to a, but not equal to it, then $g(x)$ is very close to L".

By analogy with our results for sequences we have:

Rigorous Definition of a Limit

We say the number L is the limit of the function f with domain R^+ if, for every positive number ε, there is a number T such that, for all $t > T$,

$$f(t) \in [L - \varepsilon, L + \varepsilon]$$

i.e.

$$L - \varepsilon \leqslant f(t) \leqslant L + \varepsilon$$

Similarly we have:

Rigorous Definition of a Limit

A limit of a function g near a point a is a number L such that for each positive number ε, however small, there is a positive number δ such that the set $\{x : 0 < |x - a| \leq \delta$ and $x \in$ the domain of $g\}$ is non-empty, and its image under g is a subset of $[L - \varepsilon, L + \varepsilon]$.

In Chapter 4 we also considered the idea of *continuity* of a function. The definition was:

Definition of Continuity

If f is a real function and a is an element of its domain, then "f is continuous at a" is equivalent to the statement "$\lim_{x \to a} f(x)$ exists and is equal to $f(a)$".

This definition is now completely rigorous because the concept of the limit of a function has been made rigorous.

The definition of continuity is equivalent to saying that given any positive number ε there exists a positive number δ such that

$$|f(x) - f(a)| \leq \varepsilon$$

for all x satisfying $0 \leq |x - a| \leq \delta$. Notice that this time we do not have a strict inequality ($<$) on the left: x can be equal to a.
An immediate and important consequence is that if f is continuous at a and if g is continuous at $b = f(a)$, then $g \circ f$ is continuous at a. We can express this by saying that *a continuous function of a continuous function is a continuous function.*

Exercise 1

Prove that a continuous function of a continuous function is a continuous function.

6.5 Additional Exercises

Exercise 1

The "snowflake curve" is the limit of a sequence of polygons formed as follows:

first polygon second polygon third polygon

and so on. At each stage, every line segment

in the old figure is changed to

in the new figure in such a way as always to increase the enclosed area. Calculate the limiting area enclosed, taking the area of the triangle (first polygon) as 1 unit. What can be said about the limiting length of the perimeter? (All angles are 60° or 120°.)

Exercise 2

Use the addition and multiplication rules for limits to evaluate:

(i) $\lim\limits_{k \text{ large}} \left(4 + \dfrac{1}{k}\right)$

(ii) $\lim\limits_{k \text{ large}} (2^{-k} + \pi_k)$

where π_k denotes π rounded off to k places of decimals, i.e.

$$\pi_1 = 3.1$$

$$\pi_2 = 3.14$$

$$\pi_3 = 3.142, \text{ etc.}$$

(iii) $\lim_{k \text{ large}} (\underbrace{0.333\ldots3}_{k \text{ digits}} \times \pi_k)$

(iv) $\lim_{k \text{ large}} v_k^2$

where $v_1, v_2, \ldots$ is any convergent sequence. Express your answer in terms of $\lim \underline{v}$.

(v) $\lim_{k \text{ large}} \left(2 + \frac{1}{k}\right)\left(3 + \frac{2}{k}\right)$

6.6 Answers to Exercises

Section 6.1

Exercise 1

We want to show that for any positive ε there exists a positive integer N such that

$$-\varepsilon \leqslant \frac{1}{k^2} - 0 \leqslant \varepsilon \quad (k \in Z^+ \text{ and } k > N)$$

The left-hand inequality places no restriction on N. The right-hand inequality is equivalent (since ε and k^2 are positive) to

$$\frac{1}{\varepsilon} \leqslant k^2$$

and hence to

$$\frac{1}{\sqrt{\varepsilon}} \leqslant k$$

For the definition of a limit to be satisfied therefore, we want an N large enough to ensure that, whenever k is greater than N, then it is greater than or equal to $\frac{1}{\sqrt{\varepsilon}}$. The value of N suggested in the question is large enough for this purpose.

Exercise 2

We want to show that, for any positive ε, there is a positive integer N such that

$$\tfrac{1}{3} - \varepsilon \leqslant 0.\underbrace{333\ldots3}_{k\text{ digits}} \leqslant \tfrac{1}{3} + \varepsilon \quad (k \in Z^+ \text{ and } k > N)$$

The right-hand inequality places no restriction on N, since

$$0.\underbrace{33\ldots3}_{k\text{ digits}} \leqslant \tfrac{1}{3} \leqslant \tfrac{1}{3} + \varepsilon$$

for all allowed values of k and ε. The left-hand inequality can be written

$$\tfrac{1}{3} - \varepsilon \leqslant \tfrac{1}{3} - 0.\underbrace{00\ldots0}_{k\text{ zeros}}333\ldots$$

which is equivalent to

$$0.\underbrace{00\ldots0}_{k\text{ zeros}}33\ldots \leqslant \varepsilon \quad (k \in Z^+ \text{ and } k > N)$$

We want to find an N that is large enough to ensure that if $k > N$ then the above inequality is satisfied. The suggestion in the question is to choose N so as to give ε a decimal representation of the form

$$\varepsilon = 0.\underbrace{00\ldots0}_{N\text{ zeros}}a_1a_2\ldots$$

with $a_1 \geqslant 1$.

Then if $k > N$, the decimal representation of $\frac{1}{3} - u_k$ has more consecutive zeros after the decimal point than that of ε, so that

$$0.\underbrace{00\ldots0}_{k\text{ zeros}}33\ldots \leqslant 0.\underbrace{00\ldots0}_{N\text{ zeros}}a_1a_2\ldots$$

is indeed satisfied. Thus we have verified that for any ε there does exist an element u_N beyond which all elements are in $[\frac{1}{3} - \varepsilon, \frac{1}{3} + \varepsilon]$, and so by the definition of a limit the sequence has limit $\frac{1}{3}$.

Section 6.2

Exercise 1

The sequence $\underset{\sim}{u} + \underset{\sim}{v}$ is $5\frac{1}{2}, 5\frac{1}{3}, 5\frac{1}{4}, 5\frac{1}{5}, \ldots$. The limit of $\underset{\sim}{u}$ is 3 and the limit of $\underset{\sim}{v}$ is 2. Also, from the sequence $\underset{\sim}{u} + \underset{\sim}{v}$ as written out above, we see that the limit of $\underset{\sim}{u} + \underset{\sim}{v}$ is 5. These limits are connected by the relationship

$$\lim \underset{\sim}{u} + \lim \underset{\sim}{v} = \lim (\underset{\sim}{u} + \underset{\sim}{v})$$

Exercise 2

"$\underline{u}$ is convergent" means that we can find $N \in Z^+$ such that u_k is as close to $\lim \underline{u}$ as we please for all $k > N$.

"$\underline{v}$ is convergent" means that we can find $M \in Z^+$ such that v_k is as close to $\lim \underline{v}$ as we please for all $k > M$.

Intuitively, this means that we can find $P \in Z^+$ such that $u_k + v_k$ is as close to $\lim \underline{u} + \lim \underline{v}$ as we please for all $k > P$, where P depends on N and M; that is, $\underline{u} + \underline{v}$ is convergent to $\lim \underline{u} + \lim \underline{v}$.

This also suggests the rigorous line of proof given in the text.

Exercise 3

Multiplying the two sequences together we get the sequence

$$0.66, 0.6666, 0.666666, \ldots$$

Now $\lim \underline{u} = 0.2222\ldots = \frac{2}{9}$ and $\lim \underline{v} = 3$.

Also, from the sequence that we have just obtained, we have

$$\lim (\underline{u} \times \underline{v}) = 0.6666\ldots = \tfrac{2}{3}$$

so that

$$\lim \underline{u} \times \lim \underline{v} = \lim (\underline{u} \times \underline{v})$$

since $\frac{2}{9} \times 3 = \frac{2}{3}$.

Exercise 4

We proceed in exactly the same way as in Exercise 2. We can find integers $N, M \in Z^+$ such that u_k is as close to $\lim \underline{u}$ as we please for all $k > N$, and v_k is as close to $\lim \underline{v}$ as we please for all $k > M$. Intuitively, this means that we can find $P \in Z^+$ such that $u_k \times v_k$ is as close to $(\lim \underline{u}) \times (\lim \underline{v})$ as we please for all $k > P$, where P depends on N and M; that is, $\underline{u} \times \underline{v}$ is convergent and

$$\lim (\underline{u} \times \underline{v}) = (\lim \underline{u}) \times (\lim \underline{v})$$

This is the multiplication rule for limits. It can be proved, using the rigorous definition of a limit, but we shall not give the details here.

Section 6.3

Exercise 1

$$S_k = \begin{cases} 0 \text{ if } k \text{ is even} \\ 1 \text{ if } k \text{ is odd} \end{cases}$$

and, since the sequence of partial sums $1, 0, 1, 0, \ldots$ diverges, the series also diverges.

Exercise 2

The formula given in Example 1 gives, for the kth partial sum,

$$S_k = 1 + r + r^2 + \cdots + r^{k-1} = \frac{1 - r^k}{1 - r} \qquad (r \in R, r \neq 1).$$

We are interested in the behaviour of this expression for large k. This depends on the value of r, and so there are several cases to consider.

(i) If $|r| < 1$, then $\lim\limits_{k \text{ large}} r^k = 0$, and so

$$\lim_{k \text{ large}} \frac{1 - r^k}{1 - r} = \frac{1}{1 - r}.$$

In this case the series converges and its sum is $\frac{1}{1 - r}$.

(ii) If $r = 1$, then the formula for S_k does not apply: we see that $S_k = k$, and so the series diverges.

(iii) If $|r| > 1$, then $|r^k|$ increases with k, without any bound, and so the series diverges.

(iv) If $r = -1$, we have the series

$$1 - 1 + 1 - 1 + 1 \ldots$$

which, as we have seen in the previous solution, diverges.

Section 6.4

Exercise 1

Let f be continuous at a and g be continuous at $b = f(a)$ Then, for any positive number ε, there exists a δ such that

$$|g(y) - g(b)| \leqslant \varepsilon$$

for all $0 \leqslant |y - b| \leqslant \delta$.

Also, for this same number δ, there exists a δ' such that

$$|f(x) - f(a)| \leqslant \delta$$

for all $0 \leqslant |x - a| \leqslant \delta'$.

Comparing the two inequalities involving δ and remembering that $b = f(a)$, we see that we can write $y = f(x)$, for all

$$0 \leqslant |x - a| \leqslant \delta'$$

We thus have

$$|g(y) - g(b)| = |g(f(x)) - g(f(a))| \leqslant \varepsilon$$

for all $0 \leqslant |x - a| \leqslant \delta'$.

By definition, therefore, $g \circ f : x \longmapsto g(f(x))$ is continuous at $x = a$.

Section 6.5

Exercise 1

Let a_1 be the area of the triangle ($=1$), and for each $n = 2, 3, \ldots$ let a_n be the area added at the $(n - 1)$th stage. This area is added in the form of b_n congruent triangles, each having one-third of the linear dimensions, and therefore $\frac{1}{9}$ of the area, of those added at the previous stage. Thus, the area of each triangle added at the $(n - 1)$th stage is $(\frac{1}{9})^{n-1}$, and so

$$a_n = b_n \times 9^{-(n-1)} \qquad (n = 2, 3, \ldots).$$

Now b_n, the number of triangles added at the $(n - 1)$th stage is equal to the number of line segments created at the previous stage. At the first stage, the number of line segments created is 3, and this is multiplied by 4 each stage. Thus,

$$b_n = 3 \times 4^{(n-2)} \qquad (n = 2, 3, \ldots).$$

This gives, when substituted in the previous equation,

$$a_n = \frac{1}{3} \times \left(\frac{4}{9}\right)^{(n-2)} \qquad (n = 2, 3, \ldots).$$

The total area at the $(n - 1)$th stage is

$$a_1 + a_2 + \cdots + a_n,$$

and so the limiting area can be expressed as the infinite series:

$$1 + \frac{1}{3} + \frac{1}{3} \times \frac{4}{9} + \frac{1}{3} \times \left(\frac{4}{9}\right)^2 + \cdots + \frac{1}{3} \times \left(\frac{4}{9}\right)^k + \cdots$$

which is a geometric series with common ratio $\frac{4}{9}$, excluding the first term, which is an "odd man out". The sum is therefore, by the result of Exercise 6.3.2

$$S = 1 + \frac{a}{1 - r}, \quad \text{where} \quad a = \frac{1}{3} \quad \text{and} \quad r = \frac{4}{9},$$

$$= 1\tfrac{3}{5},$$

which is thus the limiting area.

The length of the perimeter of the snowflake does not fare so happily, however. At each stage the number of line segments is multiplied by 4, and the length of each line segment is divided by 3, so that the *total* length of the perimeter is *multiplied* by $\frac{4}{3}$ at each stage. Thus the lengths of the successive polygons are

$$1, \frac{4}{3}, \left(\frac{4}{3}\right)^2, \left(\frac{4}{3}\right)^3, \ldots$$

This sequence diverges, and in fact the length increases beyond all bounds. There is no "limiting length".

Exercise 2

In (i) and (ii) below we use the addition rule

$$\lim \underset{\sim}{u} + \lim \underset{\sim}{v} = \lim (\underset{\sim}{u} + \underset{\sim}{v})$$

In (iii) and (iv) we use the multiplication rule

$$(\lim \underset{\sim}{u}) \times (\lim \underset{\sim}{v}) = \lim (\underset{\sim}{u} \times \underset{\sim}{v})$$

In (v) we combine both these rules.

(i) $$\lim_{k\,\text{large}} \left(4 + \frac{1}{k}\right) = \lim (4 + 1, 4\tfrac{1}{2}, 4\tfrac{1}{3}, \ldots)$$

$$= \lim (4, 4, 4, \ldots) + \lim (1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots)$$

$$= 4 + 0 = 4$$

(ii) $$\lim_{k\,\text{large}} (2^{-k} + \pi_k) = \lim_{k\,\text{large}} (2^{-k}) + \lim_{k\,\text{large}} (\pi_k)$$

$$= 0 + \pi = \pi$$

(iii) $$\lim_{k\,\text{large}} (0.\underbrace{33\ldots3}_{k\text{ digits}} \times \pi_k) = \lim_{k\,\text{large}} (0.\underbrace{33\ldots3}_{k\text{ digits}}) \times \lim_{k\,\text{large}} (\pi_k)$$

$$= \frac{1}{3} \times \pi = \frac{\pi}{3}$$

(iv) $$\lim_{k\,\text{large}} (v_k^2) = \lim_{k\,\text{large}} (v_k \times v_k)$$

$$= (\lim_{k\,\text{large}} v_k) \times (\lim_{k\,\text{large}} v_k) = (\lim \underset{\sim}{v})^2$$

(v) $$\lim_{k\,\text{large}} \left(2 + \frac{1}{k}\right) \times \left(3 + \frac{2}{k}\right) = \lim_{k\,\text{large}} \left(2 + \frac{1}{k}\right) \times \lim_{k\,\text{large}} \left(3 + \frac{2}{k}\right)$$

$$= 2 \times 3 = 6$$

CHAPTER 7 THE DEFINITE INTEGRAL

7.0 Introduction

In this chapter we introduce the concept of a *definite integral*. We start by looking at ways in which we evaluate area, in particular at ways in which we can find approximate values for the area of a plane surface with a curved boundary. We use the *limit* concept, introduced earlier in Chapters 4 and 6, and we arrive at our formal definition of a definite integral by considering the limits of sequence of approximations. We extend the calculation of definite integrals to include combinations of functions, and we look at the relation of the definite integral to area under the graph of a function when the graph crosses the x-axis.

7.1 Area

The history of the problem of finding area is an interesting one. In the early years of ancient Babylon it was believed that the area of a plane figure depended on its perimeter.

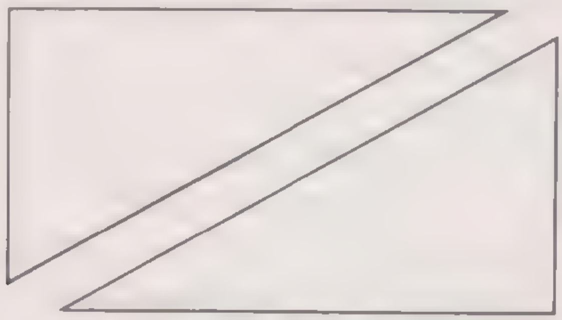
However, the correct methods for finding areas of rectangles and triangles were known before 2200 B.C. The next step, that of finding areas of plane figures with specific curved boundaries, such as a parabolic arc, was apparently not taken until the time of Archimedes (287–212 B.C.). His mathematical method is essentially the same as the one we develop in this chapter. One of his elementary checks was to cut out the appropriate shape in a material of uniform density, and compare its weight with that of a shape of the same material and of known area. Archimedes realized that the problem of the determination of volumes bounded by curved surfaces was similar to the problem of determination of area, and that both could be approached by using a process involving closer and closer approximations. Much later, in the seventeenth century, Newton, amongst others, formalized integration and established its link with differentiation. Newton's work aroused enormous interest, and the names of many mathematicians of renown appear in the history of the calculus. By the time Riemann published his definition of the definite integral in 1854, mathematicians had long realized that they had in their hands an extremely powerful tool.

Intuitively we all know what we mean by area, just as we feel we know the meanings of length, time, speed and volume. We are also aware that we use each of these words in two senses: sometimes to mean a physical quantity, and sometimes to mean a measurement of that quantity. For example, the word "area" means a vacant piece of level ground; but to avoid using a clumsy phrase, we say "the area is five acres", where

the word "area" stands for "the measurement of the area". The sense intended is usually obvious from the context. Here, we are interested in the measurement of area. We can leave the philosophers to worry about the fundamental concept. As mathematicians, we use our intuition to guide us to a precise mathematical formulation of area; we then develop and generalize the mathematical concept we have defined.

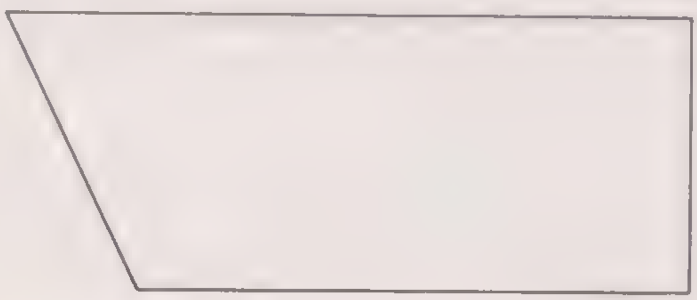
Using our intuition, let us *define* the area of a rectangle as the product of the lengths of two adjacent sides. With this as our starting point, we shall proceed to define areas of other figures of increasing complexity, always assuring ourselves that our definitions accord with our intuitive expectations.

Let us start with rectilinear figures, which are figures bounded by straight lines. By fitting together two identical right-angled triangles to form a rectangle, we find that the area of such a triangle is half the product of the length of a base and the corresponding height, which we shall write as $\frac{1}{2}$ × base × height.



Knowing this, we can find the area of any rectilinear figure by regarding the figure as being built up of rectangles and right-angled triangles.

Exercise 1



A trapezium is a quadrilateral which has two parallel sides.

Show that the area of the particular trapezium shown is $\frac{1}{2}ad$, where

a = the sum of the lengths of the parallel sides, and

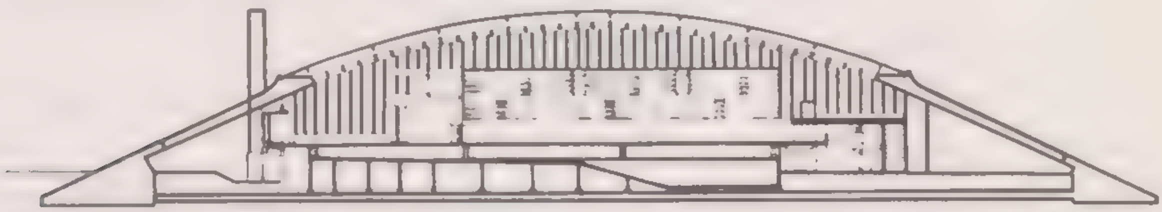
d = the perpendicular distance between them.

Difficulties arise however, when we consider the area of a region bounded by a curve such as a circle or an arch. We can, like Archimedes, cut out the appropriate shape and weigh it, and then say that the area is the area of the rectangle of the same material and weight. A mathematical definition is more complicated and is phrased in terms of limiting processes, but it is much more satisfactory.

Let us first look at a specific example:

Example 1

The Dollan Baths, Scotland's Olympic length swimming baths, were officially opened on May 27, 1968, by R. B. McGregor, the Scottish International swimmer.

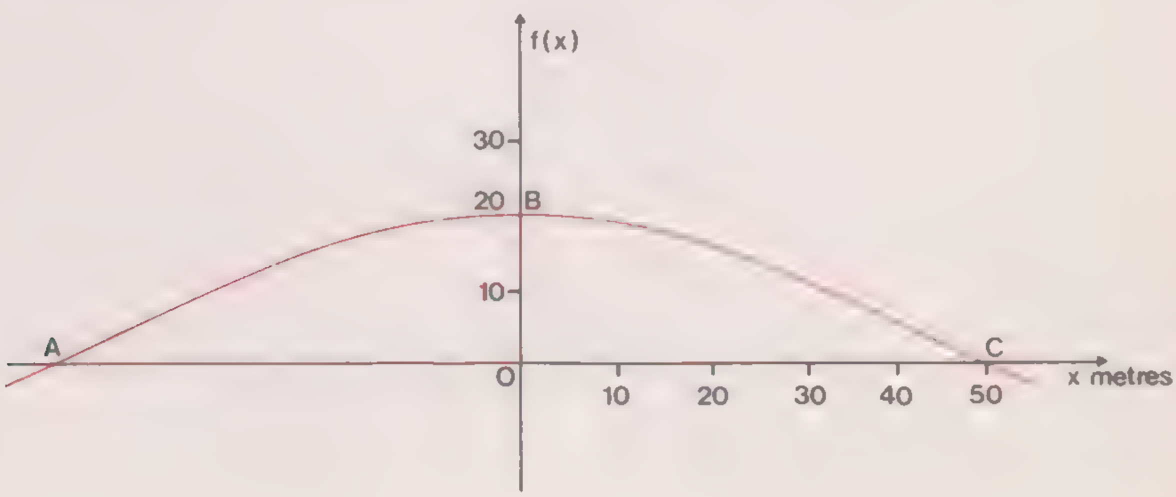


(Courtesy the Cement and Concrete Association.)

Designed by architect A. Buchanan Campbell, the building rises in one immense parabolic arch, which has a span of 100 metres and which reaches a maximum height of 20 metres above ground level.

When he had decided on the shape, the architect had to calculate the cross-sectional area of the building to enable him to find the pressure exerted on the structure.

In order to emulate his calculation, we first need to find a function f whose graph is a curve which represents the outline of the building on a suitable scale.



You may like to try to find this function, knowing that its graph is part of a parabola, which means that it is a function of the form

$$x \longmapsto a + bx + cx^2 \qquad (a, b, c \in R)$$

You should get the answer $x \longmapsto 20 - \dfrac{x^2}{125}$. However, our aim is to calculate the area underneath this graph. There are two parts to this:

(i) we approximate to the area with rectangles, find upper and lower estimates of the cross-sectional area of the parabola and hence find the best estimate of the area in the two cases:
  (a) when OC is divided into five equal intervals;
  (b) when OC is divided into ten equal intervals;

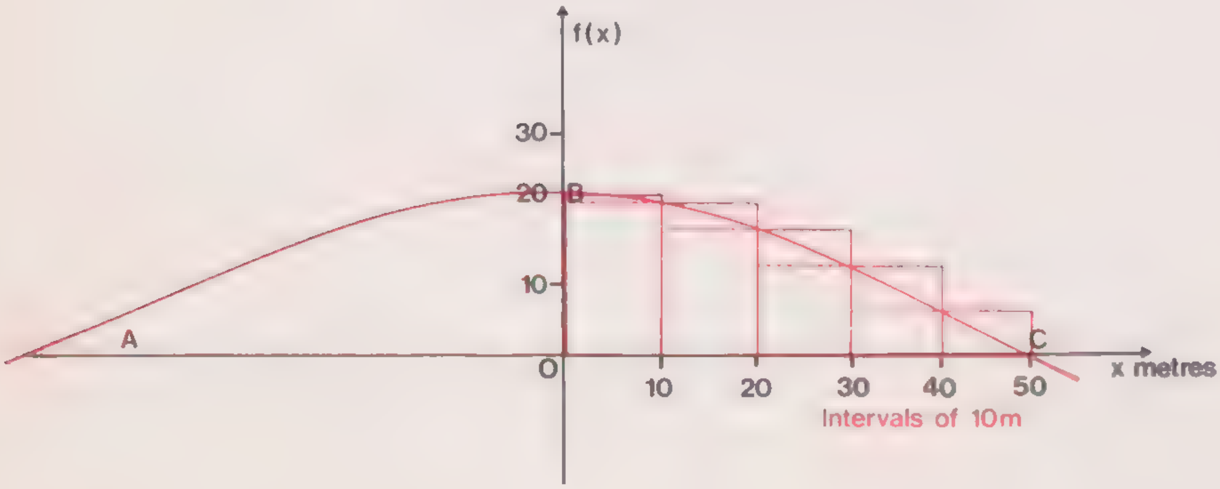
(ii) we determine how many intervals of OC would be required to ensure that the estimated error is less than $1\,m^2$.

(i) The basic graph which corresponds to the architect's cross-section is shown above.

The parabolic arch is symmetric; therefore, the area represented by $AOCB$ is twice the area represented by OBC. The images for the given values of the variable x are:

x	0	5	10	15	20	25	30	35	40	45	50
$f(x)$	20	19.8	19.2	18.2	16.8	15	12.8	10.2	7.2	3.8	0

(a) Using five intervals for the half area we now form the area consisting of rectangles which give an area larger than the area required.



These rectangles are outlined by solid lines in the figure. This gives us a first upper estimate for the *total* area:

$$2 \times 10\{20 + 19.2 + 16.8 + 12.8 + 7.2\}\ \mathrm{m}^2 = 1520\ \mathrm{m}^2$$

The dashed lines in the figure indicate an area made up of rectangles which give an area smaller than the area we require.
This gives a first lower estimate for the total area:

$$2 \times 10\{19.2 + 16.8 + 12.8 + 7.2 + 0\}\ \mathrm{m}^2 = 1120\ \mathrm{m}^2$$

We notice that the difference between the two estimates is twice the area of the largest rectangle.

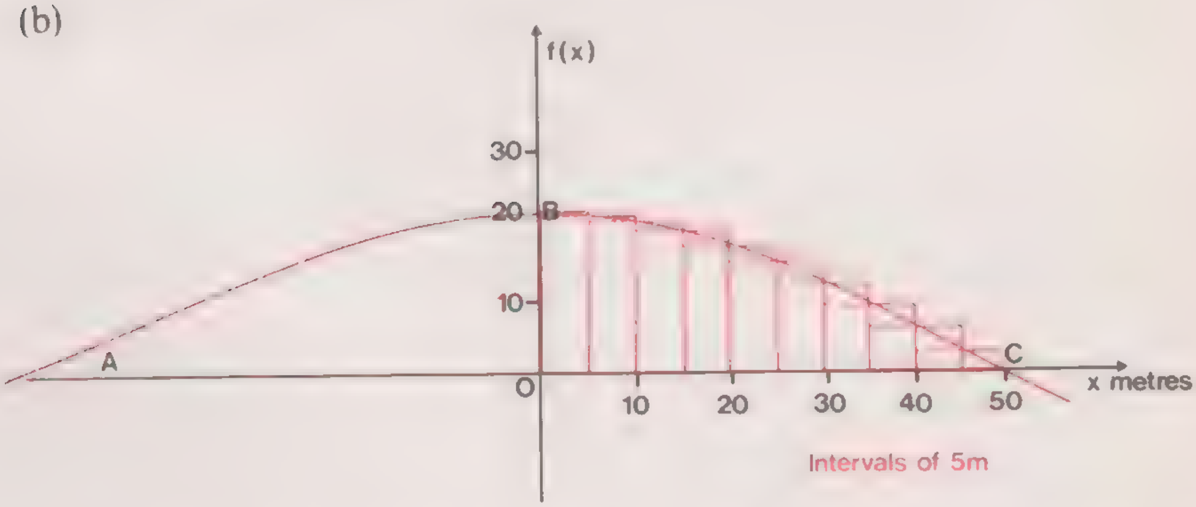
Thus the best estimate for the total area using five intervals is

$$\frac{1520 + 1120}{2}\ \mathrm{m}^2 = 1320\ \mathrm{m}^2$$

with estimated error

$$\frac{1520 - 1120}{2}\ \mathrm{m}^2 = 200\ \mathrm{m}^2$$

(b)



When ten intervals are used for the half area, we find that the upper bound on the area using ten rectangles is now

$$2 \times 5\{20 + 19.8 + \cdots + 7.2 + 3.8\}\ \mathrm{m}^2 \qquad = 1430\ \mathrm{m}^2$$

Calculation of the lower bound gives an area $= 1230\ \mathrm{m}^2$

Our estimate would be $\left(\dfrac{1430 + 1230}{2}\right)\ \mathrm{m}^2 \qquad = 1330\ \mathrm{m}^2$

with maximum error now $= 100\ \mathrm{m}^2$

(ii) We are now looking at a different aspect of the problem.

We state at the beginning how accurate we want our estimate of the total area to be. We want it to be accurate to within $\pm 1\ \text{m}^2$

For what width of interval can we achieve this? As we have already noticed, the difference between the upper and lower bounds of the estimate of the area is twice the area of the largest rectangle, and it is also twice the estimated error. Therefore the area of this rectangle must be less than or equal to $1\ \text{m}^2$. But in this example we know that the height of the largest rectangle is equal to 20 m regardless of the number of intervals. Therefore, to achieve the desired accuracy, its width must be at most 0.05 m, since

$$20 \times 0.05\ \text{m}^2 = 1\ \text{m}^2$$

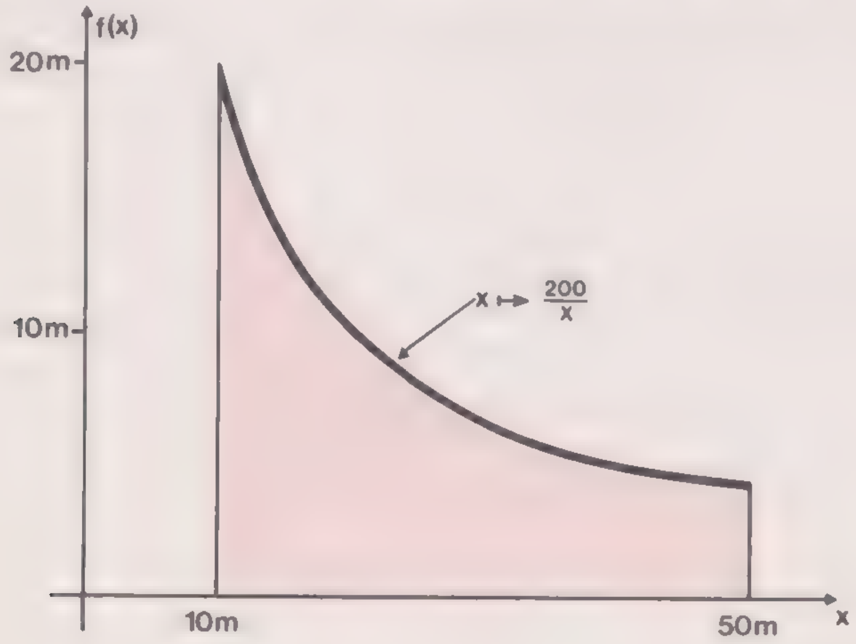
The total number of intervals along $OC = \dfrac{50\ \text{m}}{\text{interval width}} = 1000$. So to be sure of achieving the desired accuracy by this method, we must divide OC into at least 1000 intervals.

Similarly, we can show that, to achieve an estimate of the total area with maximum error $\pm 0 \cdot 1\ \text{m}^2$, we would require at least 10 000 intervals; and that in general to achieve an estimate with error $\varepsilon\ \text{m}^2$ we would require at least n sub-intervals where n is an integer such that

$$n \geqslant \frac{1000}{\varepsilon}$$

N.B. Later we shall see that we can find the above parabolic area exactly. Its value is $1333\frac{1}{3}\ \text{m}^2$. So, in fact, the above estimates are very good ones with an error which is already less than $15\ \text{m}^2$ in the five-interval case.

Exercise 2



The boundary of the cross-section of the roof of a main airport building is the graph of the function

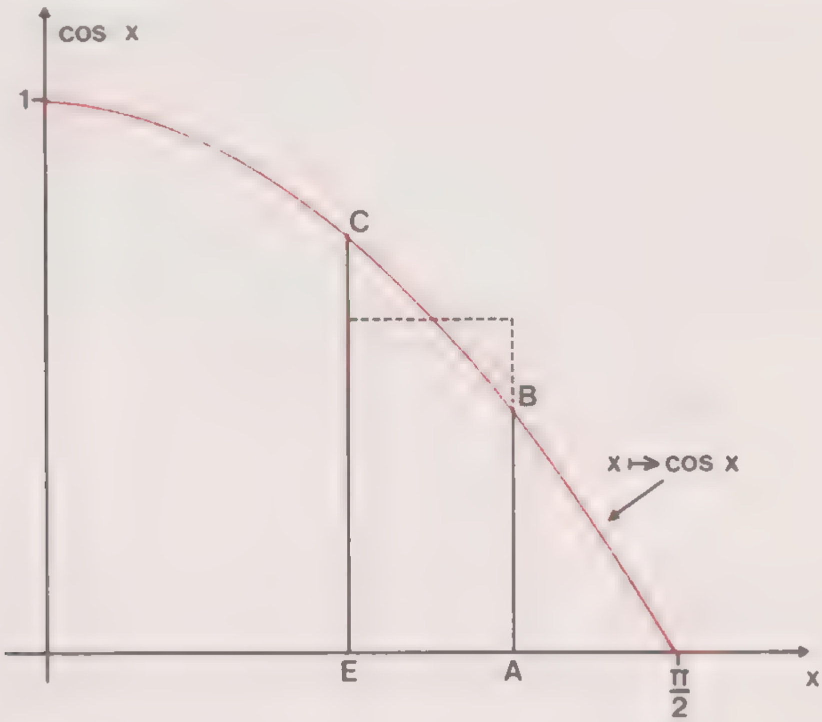
$$f: x \longmapsto \frac{200}{x} \qquad (x \in [10, 50])$$

The supporting walls are vertical and situated at $x = 10$ m and $x = 50$ m respectively.

(i) Using four intervals, estimate the cross-sectional area (indicated in the diagram) as accurately as you can by the method we have used in the example, and state the accuracy which you think you have attained.

(ii) How many intervals would you need to attain an estimate with maximum error $\pm\frac{1}{2}$ m^2?

Exercise 3



Given the function

$$f: x \longmapsto \cos x \qquad \left(x \in \left[0, \frac{\pi}{2}\right]\right)$$

we can find an estimate for the area between its graph and the two axes, by forming rectangles in the following way:
We divide the domain into equal intervals and take the height of the rectangle on each interval to be $\frac{1}{2}$ (the sum of the ordinates* of the end points of the interval). E.g. the height of the rectangle on $EA = \frac{1}{2}(AB + EC)$.

* The ordinate of a point in the Cartesian plane is the y-co-ordinate of the point.

Are the following statements true or false?

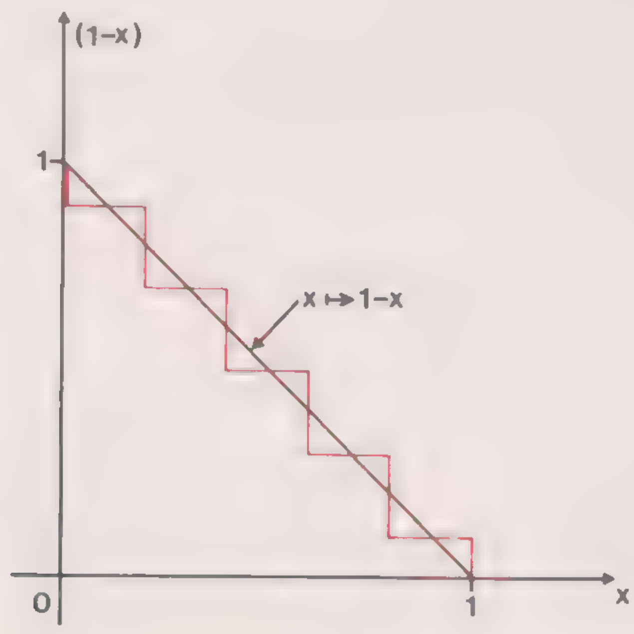
(i) The sum of the areas of the rectangles above is the same as the estimate of the area which we have used so far. That is, the estimated area is the same as that obtained by the previous method: $\frac{1}{2}$ (the sum of the areas of the larger rectangles + the sum of the areas of the smaller rectangles).

(ii) The previous method of estimating the area is better than the method described in this exercise because it gives us an estimate of the magnitude of the error.

7.2 The Definite Integral

In section 7.1 we developed a method of approximating to what we intuitively call "area", by considering a number of rectangles, this number being determined by the accuracy required, except in the cases when errors occur. In this section, we shall tie up the intituitive definition of area with a mathematical one based on the ideas you met in section 7.1. We assume that you were satisfied with the statement that, in the absence of errors, increasing the number of rectangles increases the accuracy of the estimate of the area. If so, you were quite justified, but in order to help you to see that we were making an assumption, let us consider the problem of estimating the *length* of a curve.

Exercise 1



We consider a line AB of known length to find out if an estimation procedure, similar to the one we have been considering for areas, gives us a reasonable approximation to the known length.

The graph of the function

$$x \longmapsto 1 - x \qquad (x \in [0, 1])$$

is a straight line, AB.

We approximate the length of this straight line by a "staircase", the stair treads (intervals) being of equal length. Just as in the area estimation, we can do this in a number of ways, of which we illustrate one.

(i) In the sense of finding the length of a stair carpet, find the total length of the zig-zag line in the figure above.

(ii) Does the number of steps in a staircase like the one illustrated make any difference to its total length?

(iii) Intuitively does it appear to you that we can get the corners of the staircase as near to the line AB as we like by taking enough steps?

(iv) Do you think that this implies that we can take enough steps to make the length of the staircase approximate as closely to the length of the straight line as we want?

(v) What then is the length of the line, in the limit, by this procedure?

(vi) Is this the correct length of the line? If not, what is wrong?

The purpose of this last exercise was not particularly to investigate the idea of length but to show that we must be careful in the limiting procedures we adopt to back up intuitive ideas. Approximating a curve by a "staircase" seems satisfactory when finding the area bounded by the curve, but it is unsatisfactory when trying to find the length of the curve. In both cases we are trying to generalize a concept. Whenever we do this, we must check that our generalized definition gives the answer we expect in the more fundamental case. The "staircase" approach does not give us the "right" answer for the length of a straight line: so there is no point in trying to generalize it. But our method for the calculation of area does satisfy intuition and accords with our concept of area in simple cases. Further, we were able to sandwich the required value between two estimated upper and lower bounds. Thus we can speak of the "accuracy" of our estimate, since we have trapped the numerical value of the area in an error interval which we can make as small as we please, and so we can proceed with confidence.

In this section we shall extract the mathematical definition of the definite integral which we have chosen to use. We say "chosen to use" because there are several ways of defining the definite integral, some of which are more general than others.

In section 7.3 we shall show that it is not always necessary to determine definite integrals completely from scratch.

We find that functions such as:

$$x \longmapsto 3x - x^2 + 2 \qquad (x \in R)$$

$$x \longmapsto 2 \sin x - \frac{1}{x} \qquad (x \in R^+)$$

and other similar functions can be expressed in terms of simpler functions, such as:

$$x \longmapsto 1,\ x \longmapsto x,\ x \longmapsto x^2,\ x \longmapsto x^3,\ x \longmapsto \sin x \qquad (x \in R),$$

$$x \longmapsto \frac{1}{x},\ x \longmapsto \frac{1}{x^2} \qquad (x \in R, x \neq 0)$$

Is it possible to find standard expressions for the definite integrals of simple functions such as these? If so, can we use them to find the definite integrals of more complicated functions, and thus simplify our work? The answer to the first question is "yes". We proceed to find some of these answers in this section.

The answer to the second question is also "yes" in many cases, such as the examples given above. For combinations of simple functions such as

$$x \longmapsto 3x - x^2 + 2 \qquad (x \in R)$$

we need to develop the rules for combination of the appropriate definite integrals, and this we do in section 7.3.

Using the experience we have gained in section 7.1, let us try to extract the definition of a definite integral in such a way that it matches up with our intuitive idea of area.

In section 7.1 we took a region and sandwiched it between two sets of rectangles, one containing the region and the other contained in the region. We agreed that the "area" of the region lay between two areas: the sums of the areas of each of the two sets of rectangles.

Suppose that we use this method to obtain upper and lower estimates of the area of the region by considering two sets of rectangles, each with n members.

We denote the sum of the areas of the larger rectangles by A_n and the sum of the areas of the smaller rectangles by a_n. n can be any positive integer, so if we let n take the values $1, 2, 3, \ldots$ successively, we obtain two sequences:

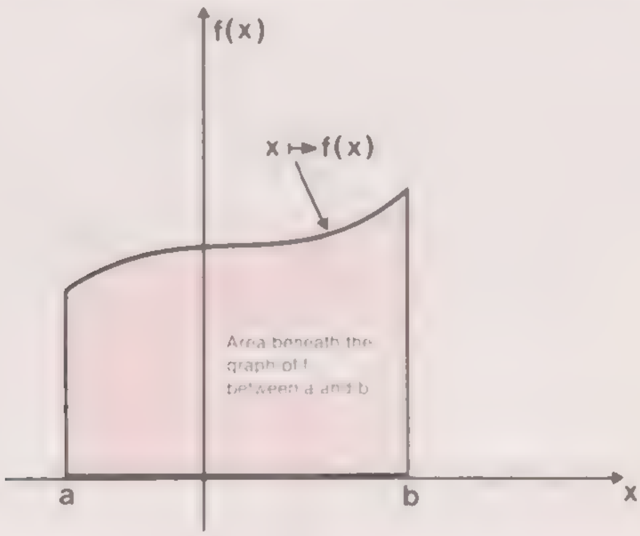
$$A_1, A_2, A_3, \ldots$$

and

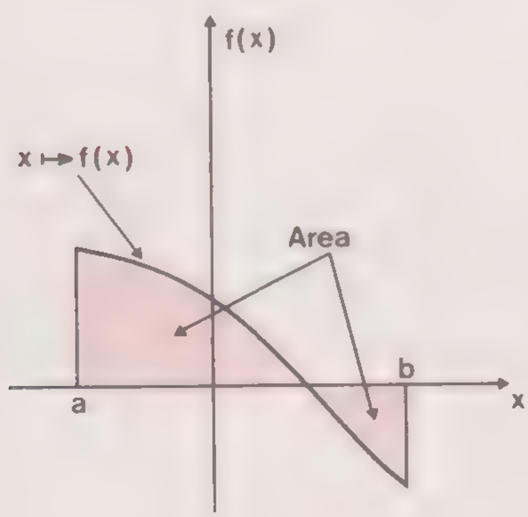
$$a_1, a_2, a_3, \ldots$$

which we denote by $\underset{\sim}{A}$ and $\underset{\sim}{a}$ respectively.

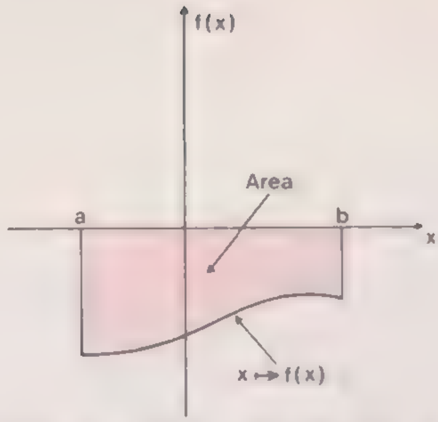
The terms of $\underset{\sim}{A}$ are always greater than the required area, and the terms of $\underset{\sim}{a}$ are always less than the required area, but as n increases, the corresponding terms in the two sequences (intuitively) get closer and closer. So as the number of rectangles gets very large, we can say (intuitively) that the two sequences have the same limit, and we can *define* this limit to be the **area** of the region. We have to be a bit more specific about area when we think in terms of more general functions. Let f be a function with domain $[a, b]$, such that the graph of f, the lines specified by $x = a$ and $x = b$, and the x-axis form the boundary of a closed region.



The area of this region is usually called the **area beneath the graph of f between a and b** even though the graph of f may look like this:



or this:

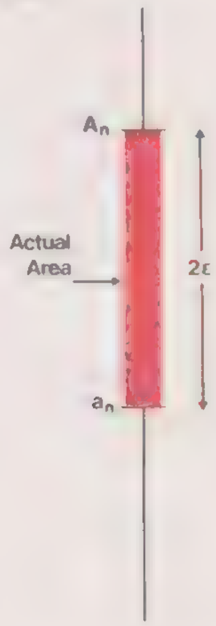


At the end of the example about the Dollan swimming baths we found that to achieve an estimate of area within the limits $\pm\varepsilon \text{ m}^2$ we would require at least n intervals, where n is an integer and

$$n \geqslant \frac{1000}{\varepsilon}$$

That is, we found we could choose n such that the estimate of the area would be within an interval of width at most 2ε, where ε could be as small as we pleased. Here we have a similar case.

Expressed the other way round, this means that as n increases the width of the error interval decreases.



For n intervals, the upper and lower bounds of the error interval are determined by A_n and a_n respectively.

If $\lim \underline{A}$ exists, then $\lim \underline{a}$ will also exist, and vice versa. The two limits will be the same number: call it A. As n increases, the width* of the error interval will shrink towards zero, and the area of the parabolic section,

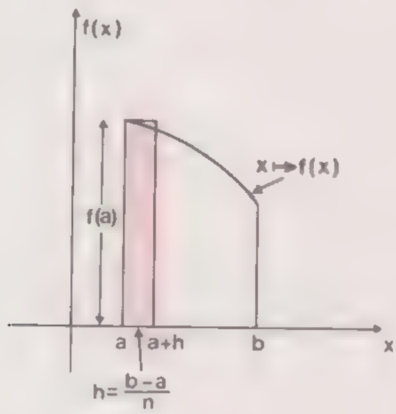
* The width of $[a, b]$ is $b - a$.

which is always contained in the error interval, must (intuitively) therefore also be A.

Now let us consider the problem for the function f more algebraically, and pick a particular sum (you will see why in a moment):

$$S_n = hf(a) + hf(a + h) + \cdots + hf(a + \{n - 1\}h)$$
$$= h[f(a) + f(a + h) + \cdots + f(a + \{n - 1\}h)]$$

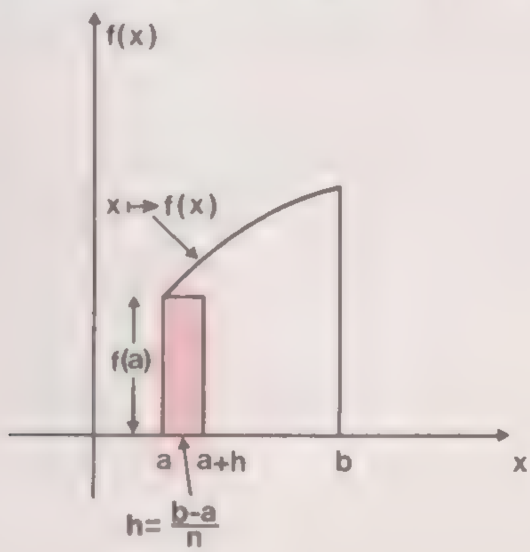
where $h = \dfrac{b - a}{n}$; that is, we have divided $[a, b]$ into n sub-intervals† each of width h. This expression for S_n defines a sequence $S_1, S_2, S_3, \ldots$ which we denote by S. For a particular function f with a graph as illustrated:



we would have

$$S_n = A_n,$$

the sum of the areas of the larger rectangles. If f has a graph like this:

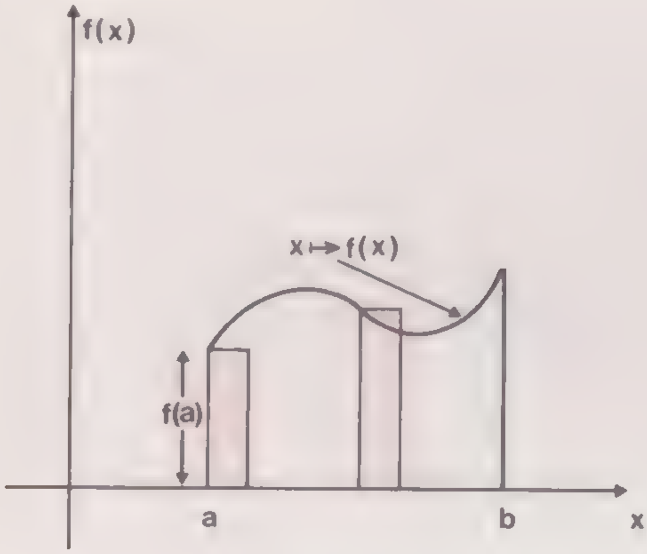


† A sub-interval of $[a, b]$ is an interval $[c, d]$ where $a \leqslant c < d \leqslant b$.

then

$$S_n = a_n$$

If f has a graph like this:



then

$$a_n < S_n < A_n$$

The point is this: if the graph of f is a continuous curve, and $f(x)$ is positive* for all values of x in $[a, b]$, then provided that

$$\lim \underset{\sim}{a} = \lim \underset{\sim}{A} = A$$

it follows that

$$\lim \underset{\sim}{S} = A$$

since S_n will be sandwiched in an error interval which diminishes in size as n gets larger (just like the estimate of the parabolic area in the Dollan Baths example).

Because A, the limit of $\underset{\sim}{S}$, has tremendously important uses in contexts other than that of area, we give it a special name. We call it the *definite integral of the function f between a and b* (or in $[a, b]$), and a special symbol

$$\int_a^b f$$

* The reason for requiring the images to be positive at this stage will become clear from a consideration of Exercise 2 which follows.

so that

$$\lim \underset{\sim}{S} = \int_a^b f$$

The symbol $\int$, an elongated "s", represents the summation process with the "a" showing where the sum begins and the "b" showing where it finishes. (The "a" and "b" are called the end-points* of the integral).

We can now forget about the rectangles, and concentrate on this particular sequence of sums, $\underset{\sim}{S}$, and its limit. We do this because we are interested not only in area but also in the importance of the definite integral in mathematics and applications. Sometimes the integral will represent an area, but on other occasions it may represent a volume, or an average, or an electric current, or a probability, or power, or distance, or a similar quantity. In each case, the definite integral will have to be interpreted with caution. To avoid some pitfalls we shall stipulate one condition on f which *guarantees* the existence of $\int_a^b f$. Remember that not all sequences converge (i.e. have a limit) and we have defined the definite integral to be the limit of a sequence. We do not want to have to return to the definition repeatedly to check on the existence of integrals. So we ask : Are there any special types of function f for which $\int_a^b f$ always exists? The answer is "Yes", and in fact we are already on solid ground. In our definition on page 150 we have included the phrase "if the graph of f is a continuous curve"; that is, the graph of f must have no "gaps" in $[a, b]$. This is the same as saying that "the function f must be continuous in $[a, b]$". It can be shown (by a more rigorous treatment than ours) that *if f is continuous in $[a, b]$*, then the definite integral $\int_a^b f$ *automatically* exists (i.e. its existence is *guaranteed*). We ask you to accept this fact. (It means that the words "provided that" before "$\lim \underset{\sim}{a} = \lim \underset{\sim}{A} = A$" on page 150 are redundant and may go home.) Hereafter, in this chapter, we shall assume that the definite integral exists (i.e. the sequence $\underset{\sim}{S}$ converges) in all our discussions.

* Some authors use the word limits.

Summary

Given an interval $[a, b]$ which is divided into n equal sub-intervals of width

$$h = \frac{b - a}{n}$$

and a function f which is continuous in $[a, b]$, we define the definite integral of f in $[a, b]$ as the limit of the sequence $\underset{\sim}{S}$, where

$$S_n = h[f(a) + f(a + h) + \cdots + f(a + \{n - 1\}h)]$$

We write

$$\lim \underset{\sim}{S} = \int_a^b f$$

Exercise 2

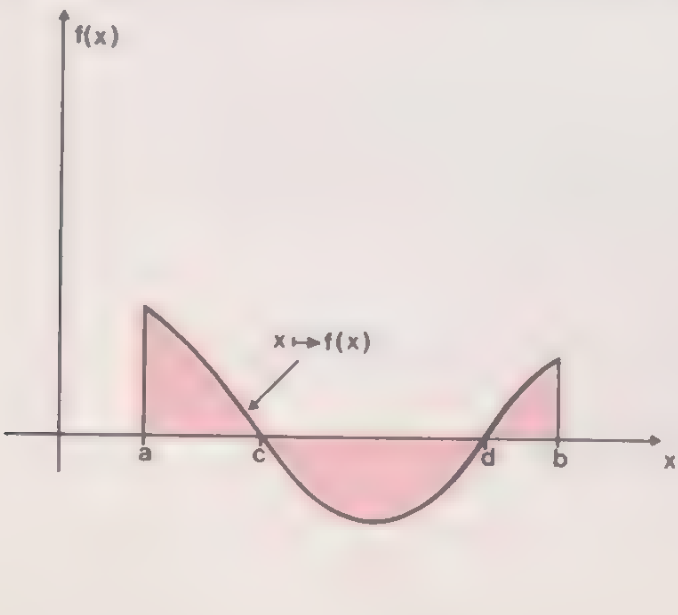
Is "the definite integral of f between a and b" synonymous with "the area beneath the graph of f between a and b"?

HINT: Consider the two definitions for the diagrams on page 147.

From the solution to Exercise 2, we may conclude that as long as we are careful to ascertain exactly where the curve representing f lies, we can use the definite integral to find the area under the graph of f between a and b. If $f(x)$ is not positive for all $x \in [c, d]$, where $[c, d]$ is a sub-interval of $[a, b]$ (see diagram), then we find the definite integral of f in the intervals

$$[a, c], \qquad [c, d], \qquad [d, b]$$

separately, and adjust the sign of the definite integral of f between c and d before adding the three results to find the total area.



The following exercise is designed to give you practice in the ideas introduced in this section and to check that they are in accordance with intuition.

Exercise 3

Find the definite integral of the function f in $[a, b]$, where

$$f: x \longmapsto 1 \qquad (x \in [a, b],\ b > a > 0)$$

and check whether it does give the area beneath the graph of f.

Exercise 4

Find the definite integral of the function f in $[a, b]$, where

$$f: x \longmapsto x \qquad (x \in [a, b],\ b > a > 0)$$

and check whether it does give the area beneath the graph of f between a and b.

HINT: In the solution you will need the fact that the sum of the first n natural numbers is

$$S_1(n) = \frac{n(n+1)}{2}$$

Just a word on notation. We have written the definite integral of a function f in $[a, b]$ as

$$\int_a^b f$$

In particular cases when f is known, for example:

$$f: x \longmapsto x \qquad (x \in [a, b])$$

we write

$$\int_a^b x \longmapsto x, \quad \text{or} \quad \int_a^b (x \longmapsto x)$$

and omit the domain of f because the part of the domain we are interested in is clear from the "a" and "b" in the notation.

In many text books you will find the notation

$$\int_a^b f(x)\,dx$$

or, in particular,

$$\int_a^b x\,dx$$

This is a notation which has been used for many years. If you are interested in its origins, you will find it in a book on the history of the subject.

If you have some familiarity with the calculus you may also wonder why here, and in the later chapters on integration and differentiation, we use a notation which is different from the classical notation with which you are familiar, and which is used in most textbooks. (In fact, you will find that there are a number of recent textbooks which use the same notation as ours.) There are several reasons for this change. The notation we use is consistent with our basic approach to mathematics through the concept of a function. Also, we wish to avoid some of the conceptual difficulties experienced by many students, which may in part be due to the traditional notation. If you are new to the subject there should be no notational difficulty. If you have already studied calculus, we suggest that you make a fresh start: the new notation should enable you to concentrate on the principles (as opposed to the techniques), because of its very novelty. Once you have mastered the basic principles, there is no objection to your using the traditional notation: the conversion from one notation to the other should not prove difficult. The definite integral we have used is not the most general one. We have, for example, chosen the particular case where we divide $[a, b]$ into equal intervals. The definite integral of f in $[a, b]$ can be satisfactorily defined in a more general way, omitting the conditions: "f is continuous in $[a, b]$" and "the sub-intervals of $[a, b]$ are of equal width". On the whole we have relied more on intuition rather than mathematical rigour.

In the last two exercises you found that

$$\int_a^b (x \longmapsto 1) = b - a \qquad \text{and} \qquad \int_a^b (x \longmapsto x) = \frac{b^2 - a^2}{2}$$

$$b > a > 0$$

By a similar method to that used in the exercises, it can be shown that

$$\int_a^b (x \longmapsto x^2) = \frac{b^3 - a^3}{3}$$

To derive this result, one needs the formula for the sum of the squares of the first n natural numbers

$$S_2(n) = \frac{n(n+1)(2n+1)}{6}$$

Exercise 5

(i) From the results so far obtained, an intuitive guess at the value of

$$\int_a^b (x \longmapsto x^m) \qquad b > a > 0, \quad m \in Z^+$$

might be

(a) $\dfrac{b^{m+1} - a^{m+1}}{m}$ (c) $\dfrac{b^{m+1} - a^{m+1}}{m+1}$

(b) $\dfrac{b^m - a^{m+1}}{m}$ (d) $\dfrac{b^m - a^m}{m}$

Which do you think is correct?

(ii) To obtain the above value would you need to know the sum of the

(a) $(m-1)$th

(b) mth

(c) $(m+1)$th

powers of the first n natural numbers?

You have probably recognized that the calculation of

$$\int_a^b (x \longmapsto x^m) \qquad (m \in Z^+)$$

from first principles becomes more cumbersome as m increases in magnitude. That is why we leave it until we get some tools which can handle it very simply. The formula we obtain if we did the calculation is, as already noted,

$$\int_a^b (x \longmapsto x^m) = \frac{b^{m+1} - a^{m+1}}{m+1} \qquad (m \in Z^+)$$

In fact, although we cannot show it at this stage, this formula is true for

$(m \in R, m \neq -1)$, and you may assume this for any subsequent exercises.

Table of Definite Integrals of Simple Polynomial Functions

f	$\int_a^b f$
$x \longmapsto 1$	$b - a$
$x \longmapsto x$	$\dfrac{b^2 - a^2}{2}$
$x \longmapsto x^2$	$\dfrac{b^3 - a^3}{3}$
$x \longmapsto x^m \; (m \in R, m \neq -1)$	$\dfrac{b^{m+1} - a^{m+1}}{m+1}$

Exercise 6

Calculate the values of

(i) $\displaystyle\int_0^4 (x \longmapsto x^3)$

(ii) $\displaystyle\int_{-2}^2 (x \longmapsto x^4)$

(iii) $\displaystyle\int_0^1 (x \longmapsto x^{43})$

7.3 The Definite Integral for Combinations of Functions

In order to extend the class of functions for which we can calculate definite integrals to include, for example,

$$\int_1^2 x \longmapsto (3x^3 + 5x)$$

we need two theorems (unless, of course, we resort to summing series!). We derive one of them in the following example and the other in the exercise below.

Example 1

Show that

$$\int_a^b cf = c\int_a^b f$$

where f is a function, c is any number and cf is the function

$$x \longmapsto cf(x)$$

We shall not give a fully rigorous proof but the essence of the argument of such a proof is here.

To define the definite integral of f between a and b we used the limit of the sequence $\underset{\sim}{S}$, where

$$S_n = h[f(a) + f(a + h) + \cdots + f(a + \{n - 1\}h)]$$

Similarly, to define the definite integral of (cf) between a and b we would use the limit of the sequence $\underset{\sim}{T}$, where

$$T_n = h[cf(a) + cf(a + h) + \cdots + cf(a + \{n - 1\}h)]$$

$$= cS_n$$

It follows that

$$\lim \underset{\sim}{T} = c \lim \underset{\sim}{S}$$

and hence

$$\int_a^b cf = c\int_a^b f$$

Exercise 1

Convince yourself, by using the definition of the definite integral, that if $f(x)$ and $g(x)$ are positive for all values of x in $[a, b]$, then

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g$$

Example 2

Calculate

$$\int_1^3 x \longmapsto (3x^3 + 5x)$$

We have

$$\int_1^3 x \longmapsto (3x^3 + 5x)$$

$$= \int_1^3 (x \longmapsto 3x^3) + \int_1^3 (x \longmapsto 5x), \text{ from rule in Exercise 1}$$

$$= 3\int_1^3 (x \longmapsto x^3) + 5\int_1^3 (x \longmapsto x), \text{ from rule in Example 1}$$

$$= 3\left(\frac{3^4 - 1^4}{4}\right) + 5\left(\frac{3^2 - 1^2}{2}\right)$$

$$= 80$$

Exercise 2

Calculate

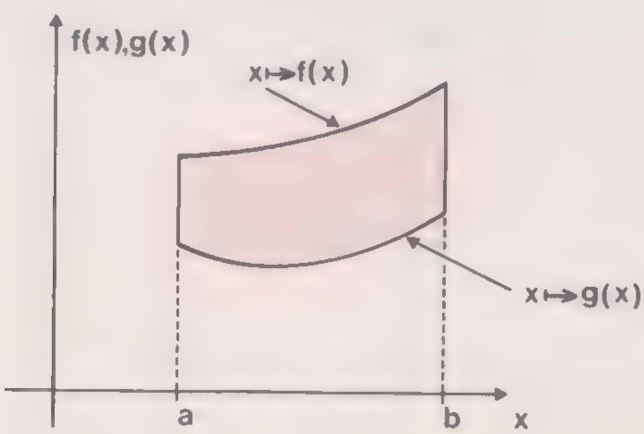
(i) $\int_2^4 x \longmapsto (2x^2 + 7x - 3)$ (iii) $\int_2^3 x \longmapsto (x^2 - x)$

(ii) $\int_{-2}^2 x \longmapsto (4 - x^2)$ (iv) $\int_0^3 x \longmapsto (x - 1)(x - 2)$

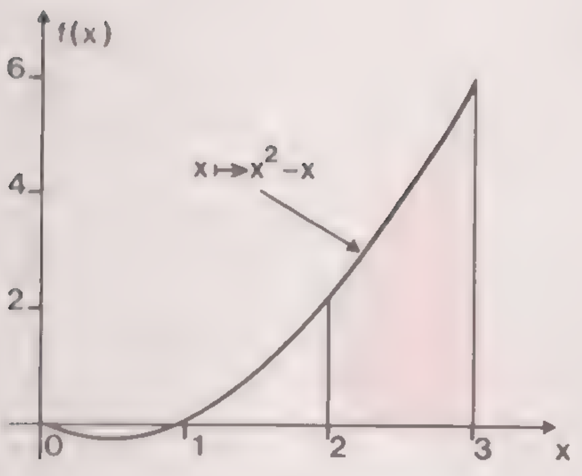
(v) What is the area beneath the graph of f in (iv)?

It is frequently possible to find areas of more complex regions with very little extra effort.

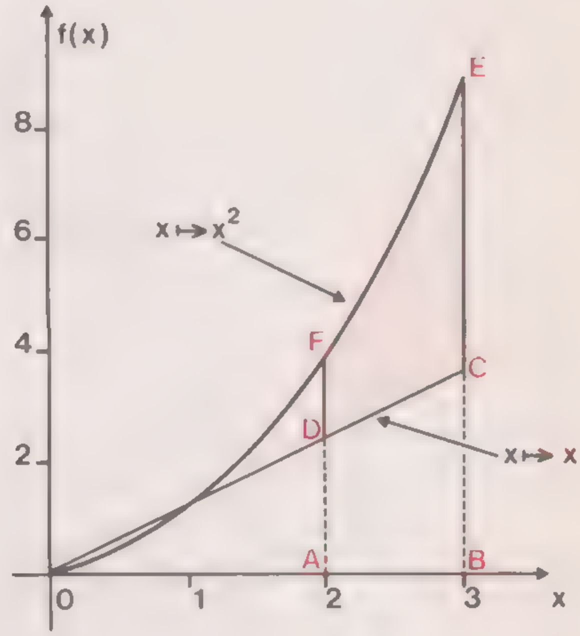
Suppose we wish to find areas which are bounded by curves and lines specified by equations of the form $x = c$, $c \in R$, but are not bounded by the x-axis; for example, the area shaded in red in the diagram.



Rather than try to form rectangles to determine the numerical value for this area, we can consider it as the difference of two areas bounded by the x-axis. Look more closely at part (iii) of the last exercise from a graphical point of view.



(a)



(b)

Since the images of x under the functions $x \longmapsto (x^2 - x)$, $x \longmapsto x^2$, $x \longmapsto x$ are not negative for x in $[2, 3]$, the definite integral

$$\int_2^3 x \longmapsto (x^2 - x)$$

represents the area shaded in red in diagram (a), and the definite integrals

$$\int_2^3 (x \longmapsto x^2), \qquad \int_2^3 (x \longmapsto x)$$

represent the areas $ABEF$ and $ABCD$ respectively in diagram (b).

The shaded areas in (a) and (b) may not *look* equal, but since

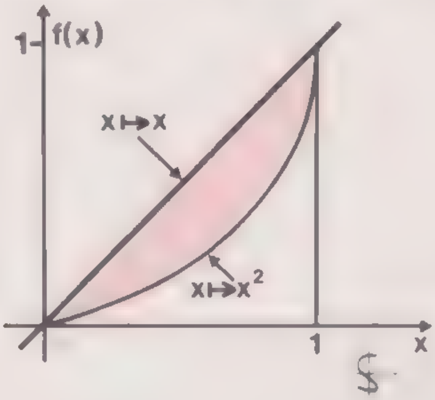
$$\int_2^3 (x \longmapsto (x^2 - x)) = \int_2^3 (x \longmapsto x^2) - \int_2^3 (x \longmapsto x)$$

these areas are the same.

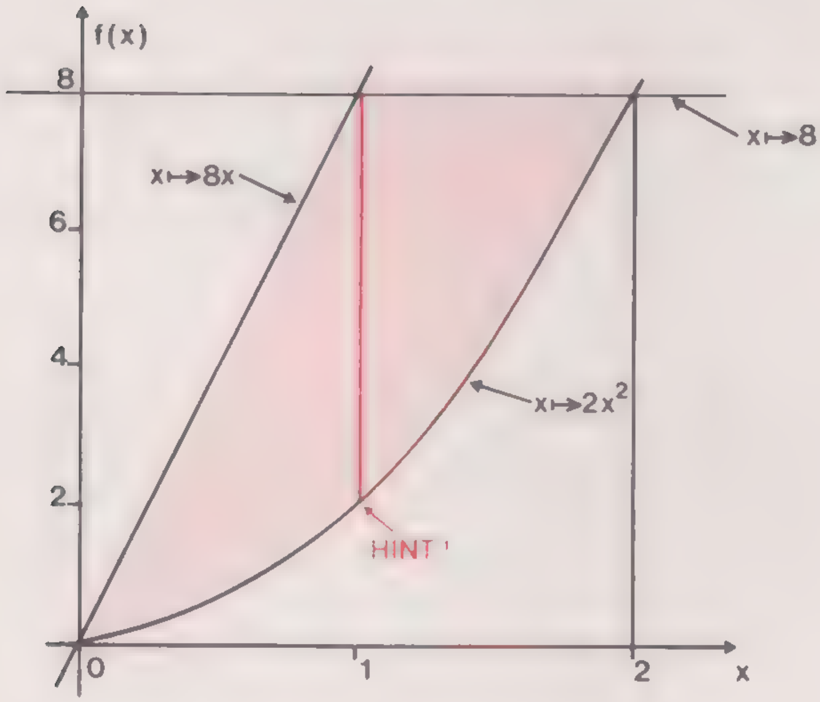
Exercise 3

Write down the definite integral, or sum of definite integrals, which represents the areas on the following graphs, and hence evaluate them.

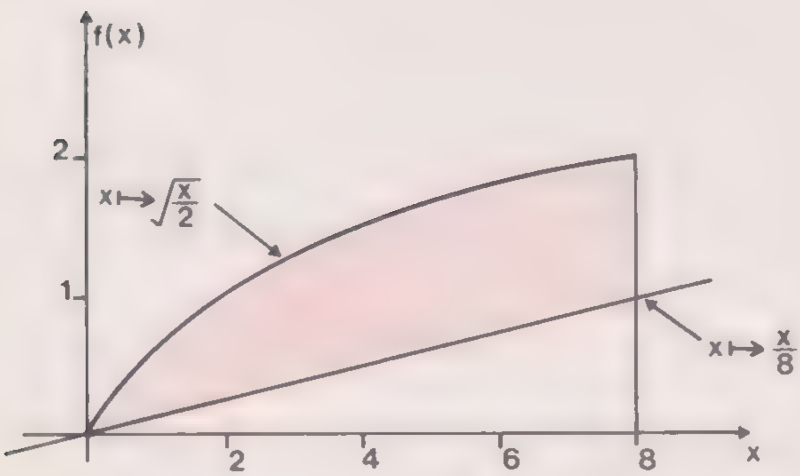
(i)



(ii)



(iii)



For this one, you will need the result

$$\int_a^b x \longmapsto x^{1/2} = \tfrac{2}{3}(b^{3/2} - a^{3/2})$$

(iv) Is it a coincidence that (ii) and (iii) have the same answer?

7.4 Additional Exercises

Exercise 1

Calculate

(i) $\displaystyle\int_0^2 x \longmapsto x^2 - 2x - 1 \qquad (x \in R)$

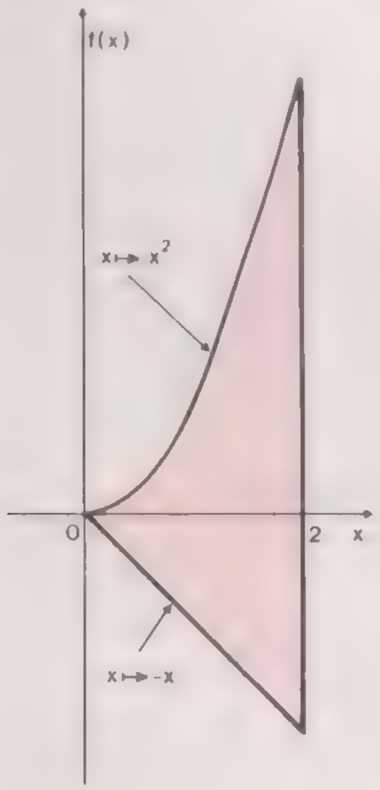
(ii) $\displaystyle\int_0^2 x \longmapsto |x^2 - 2x - 1| \qquad (x \in R)$

(iii) $\displaystyle\int_0^2 x \longmapsto |x^2 - 2x| - 1 \qquad (x \in R)$

Sketch the graph in each case.

Exercise 2

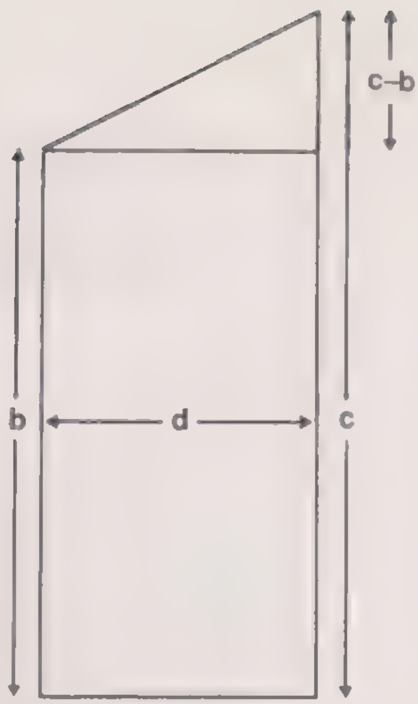
If the shaded area represents a piece of metal 1 cm thick, and x and its images are measured in centimetres, calculate the volume of material in such a metal plate.



7.5 Answers to Exercises

Section 7.1

Exercise 1



In the diagram,

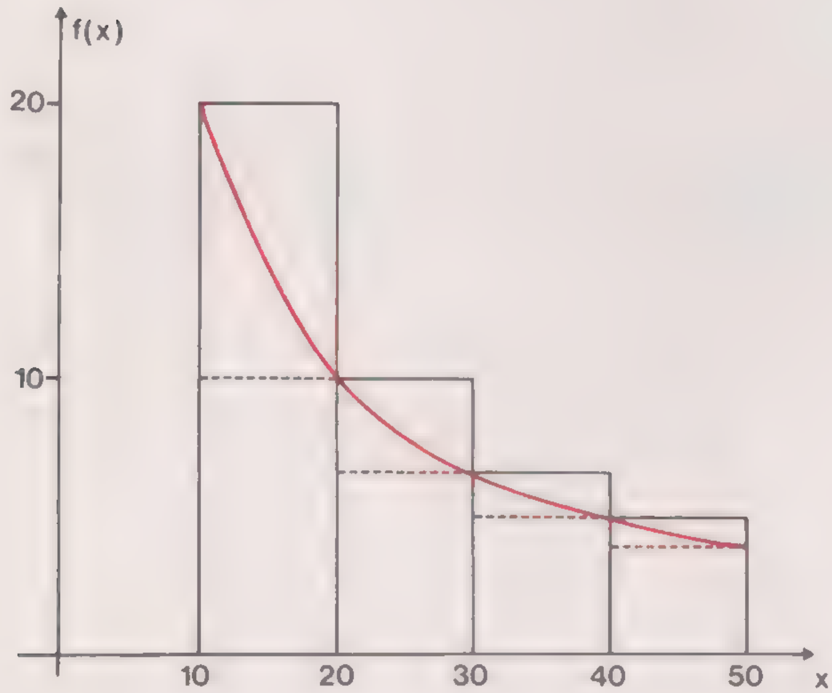
the area of the triangle $= \frac{1}{2}d(c - b)$

the area of the rectangle $= db$

So

$$\begin{aligned}\text{the area of the trapezium} &= db + \tfrac{1}{2}dc - \tfrac{1}{2}db \\ &= \tfrac{1}{2}(c + b)d \\ &= \tfrac{1}{2}ad\end{aligned}$$

Exercise 2



(i)

x	10	20	30	40	50
$f(x)$	20	10	6.67	5	4

The total area of the larger rectangles $= 416.7\,\text{m}^2$.

(Notice that in this exercise an inherent error, due to the finite decimal

representation of $\frac{2}{3}$, has been introduced. However, we can always make this as small as we please.)

The total area of the smaller rectangles $= 256.7\ \text{m}^2$. The estimate for the required area is therefore $\dfrac{(416.7 + 256.7)}{2}\ \text{m}^2 = 336.7\ \text{m}^2$, with

estimated error $= \dfrac{160}{2}\ \text{m}^2 = 80\ \text{m}^2$.

(ii) The differences in height and area between the largest and smallest rectangles are 16 m and $16 \times$ (interval width) m^2 respectively, so we require $16 \times$ (interval width) $= 1$. Therefore the width of each rectangle must be $\frac{1}{16}$ m. The total number of intervals required is therefore $\dfrac{40}{\frac{1}{16}} = 640$.

We noticed previously in Example 1 that the difference between the sum of the larger rectangles and the sum of the smaller rectangles is a very crude estimate of the error. In fact, the exact area in this case can be calculated and it is $321.9\ \text{m}^2$ (to one place of decimals), which lies well inside the error interval which we obtained in (i).

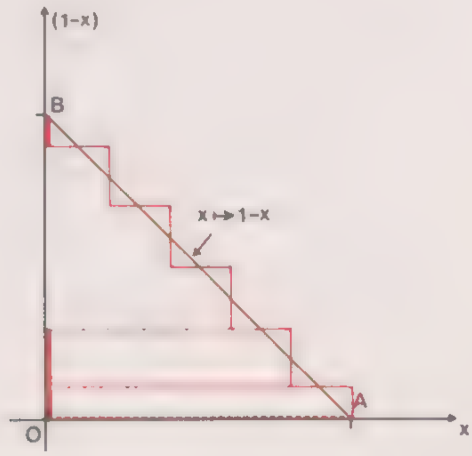
Exercise 3

(i) TRUE. When in our original method we take the estimate $\frac{1}{2}$ (the sum of the areas of the larger rectangles + the sum of the areas of the smaller rectangles) this is exactly the same as the sum of $\frac{1}{2}$ (the area of the larger rectangle + the area of the smaller rectangle) taken over all the intervals.

(ii) TRUE. The original method *also* told us the magnitude of the error we might be making. Although this error estimate may be a wild exaggeration, it is better to have some estimate than none at all.

Section 7.2

Exercise 1



(i) Each rising tread is equal to a corresponding length on the $(1 - x)$-axis. These lengths do not overlap, and therefore they add up to the length OB.

Each step length is equal to a corresponding length on the x-axis, and again the lengths do not overlap, but add up to the length OA. So

the total length of the zig-zag line = the total length of the carpet required

$= 1 + 1 = 2$

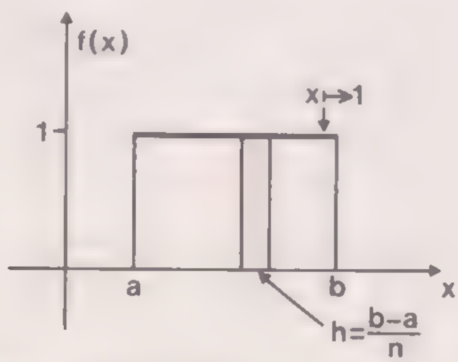
(ii) NO. The total length remains unchanged no matter how many steps are used to cover the line.
(iii) Intuitively YES.
(iv) You have probably answered YES.
(v) Apparently 2 units, using (i) and (ii).
(vi) NO. The correct length is $\sqrt{2}$ units. The stair-carpet, or zig-zag line, is always either horizontal or vertical and is never lying flat on the slope of the hypotenuse of the triangle. So, although the answer to (iii) is correct, it is misleading.

Exercise 2

Your solution should contain the following points:

(i) Even if the curve is wholly above the x-axis, if it goes up *and* down then the two sequences $\underset{\sim}{A}$ and $\underset{\sim}{S}$ are not the same, but, intuitively, the terms of the sequence $\underset{\sim}{S}$ lie between the terms of the two sequences $\underset{\sim}{a}$ and $\underset{\sim}{A}$. If these latter two sequences have the same limit, then so also has $\underset{\sim}{S}$ and we may therefore give an intuitive YES.
(ii) If some of the curve lies below the x-axis, then we have negative terms in $\underset{\sim}{S}$, hence lim $\underset{\sim}{S}$ is definitely *not* the area. So the definite integral and the area may both exist and be different.

Exercise 3



The interval is divided into n sub-intervals of width h, where $nh = b - a$.

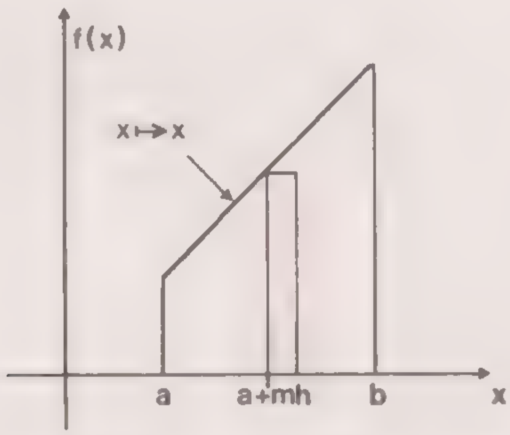
Then

$$S_n = h[f(a) + f(a+h) + \cdots + f(a + \{n-1\}h)]$$
$$= \underset{(n \text{ terms})}{h[1 + 1 + \cdots + 1]} = nh = b - a$$

The definite integral between a and b is the limit of the sequence $\underset{\sim}{S}$, that is $(b - a)$.

This is clearly the area beneath the graph between a and b.

Exercise 4



The interval is divided into n sub-intervals of width h, where $nh = b - a$. We have $f(x) = x$, so that

$$f(a) = a, f(a+h) = a + h, \ldots, f(a + \{n-1\}h) = a + \{n-1\}h$$

Therefore

$$S_n = h[f(a) + f(a+h) + f(a+2h) + \cdots + f(a + \{n-1\}h)]$$
$$= h[a + (a+h) + (a+2h) + \cdots + (a + \{n-1\}h)]$$
$$= nha + h^2[1 + 2 + 3 + \cdots + \{n-1\}]$$

(collecting together like terms)

$$= a(b-a) + \frac{(b-a)^2}{n^2}[1 + 2 + 3 + \cdots + \{n-1\}]$$

(substituting for h)

$$= a(b-a) + \frac{(b-a)^2}{n^2} S_1(n-1)$$

$$= a(b-a) + \frac{(b-a)^2}{n^2}\frac{(n-1)n}{2} \quad \text{(substituting for } S_1(n-1))$$

$$= a(b-a) + \frac{(b-a)^2}{2}\left(1 - \frac{1}{n}\right)$$

The limit of this sequence for n large is

$$a(b-a)+\frac{(b-a)^2}{2}=\frac{(b^2-a^2)}{2} \qquad \left(\text{since } \lim_{n \text{ large}}\left(1-\frac{1}{n}\right)=1\right)$$

So the definite integral of f in $[a, b]$ is $\frac{1}{2}(b^2-a^2)$.

By regarding the area beneath the graph of f between a and b as the difference between the areas of the larger triangle, whose base is the x-axis from the origin to b, and the smaller triangle, whose base is the x-axis from the origin to a, we see that this area is

$$\tfrac{1}{2}b^2-\tfrac{1}{2}a^2$$

which is the same as the definite integral of f in $[a, b]$.

Exercise 5

(i) (c) (ii) (b)

Exercise 6

(i) 64 (ii) $\frac{64}{5}$ (iii) $\frac{1}{44}$

Section 7.3

Exercise 1

Write $S_n = h[f(a)+f(a+h)+\cdots+f(a+\{n-1\}h)]$

$T_n = h[g(a)+g(a+h)+\cdots+g(a+\{n-1\}h)]$

and take the appropriate limits.

Exercise 2

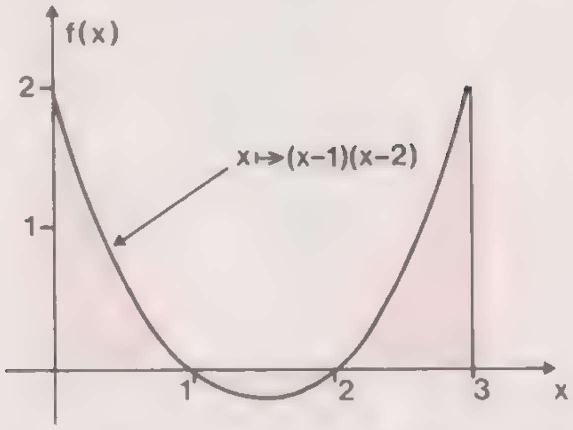
(i) $\displaystyle\int_2^4 (x \longmapsto (2x^2+7x-3))$

$$=2\int_2^4 (x\longmapsto x^2)+7\int_2^4 (x\longmapsto x)-3\int_2^4 (x\longmapsto 1)$$

$$=2\left(\frac{4^3-2^3}{3}\right)+7\left(\frac{4^2-2^2}{2}\right)-3(4-2)$$

$$=\frac{220}{3}$$

(ii) $\frac{32}{3}$
(iii) $\frac{23}{6}$
(iv) $\frac{3}{2}$
(v)



The graph of the function is shown in the diagram, and because part of the graph of this function lies below the x-axis, the definite integral evaluated in (iv) does not give the required area. So we divide the definite integral into 3 separate limiting sequences. In the first and third sequence all the terms are positive whilst in the second they are negative. The appropriate definite integrals are

$$\int_0^1 (x \longmapsto (x^2 - 3x + 2)) = \tfrac{5}{6}$$

$$\int_1^2 (x \longmapsto (x^2 - 3x + 2)) = -\tfrac{1}{6}$$

$$\int_2^3 x \longmapsto (x^2 - 3x + 2) = \tfrac{5}{6}$$

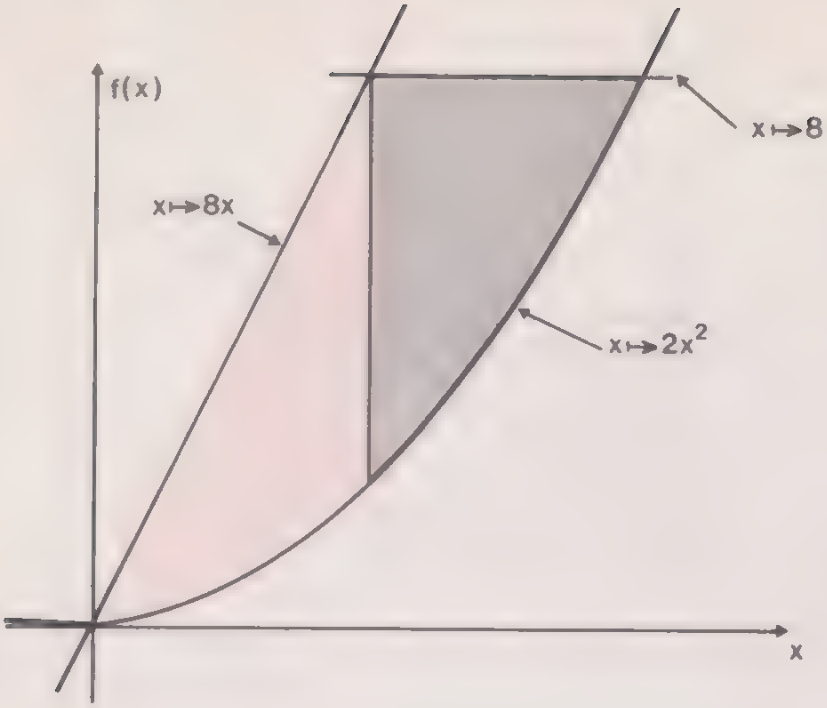
The *area* is the sum of the *magnitudes* of these three definite integrals $= \frac{5}{6} + \frac{1}{6} + \frac{5}{6} = \frac{11}{6}$.

Exercise 3

(i) $\displaystyle\int_0^1 x \longmapsto (x - x^2) = \tfrac{1}{6}$

(ii) $\displaystyle\int_0^1 x \longmapsto (8x - 2x^2) + \int_1^2 x \longmapsto (8 - 2x^2) = \tfrac{20}{3}$

The first definite integral represents the area shaded in red, the second the area shaded in black. First find the appropriate intersections of the graphs, then write down the integrals to give the result above.



(iii) $$\int_0^8 x \longmapsto \left(\frac{1}{\sqrt{2}} \times x^{1/2} - \frac{x}{8}\right) = \frac{20}{3}$$

(iv) NO. In (iii) we have drawn the graphs of the inverses of the functions used in (ii). Effectively we could have used the same diagram as in (ii) and interchanged the axes. We often face the choice of forming the definite integral "the other way", that is, using the inverse functions. The simplicity gained in the representation of the area can, however, be lost in the awkwardness of the new function, as we see in this example.

Section 7.4

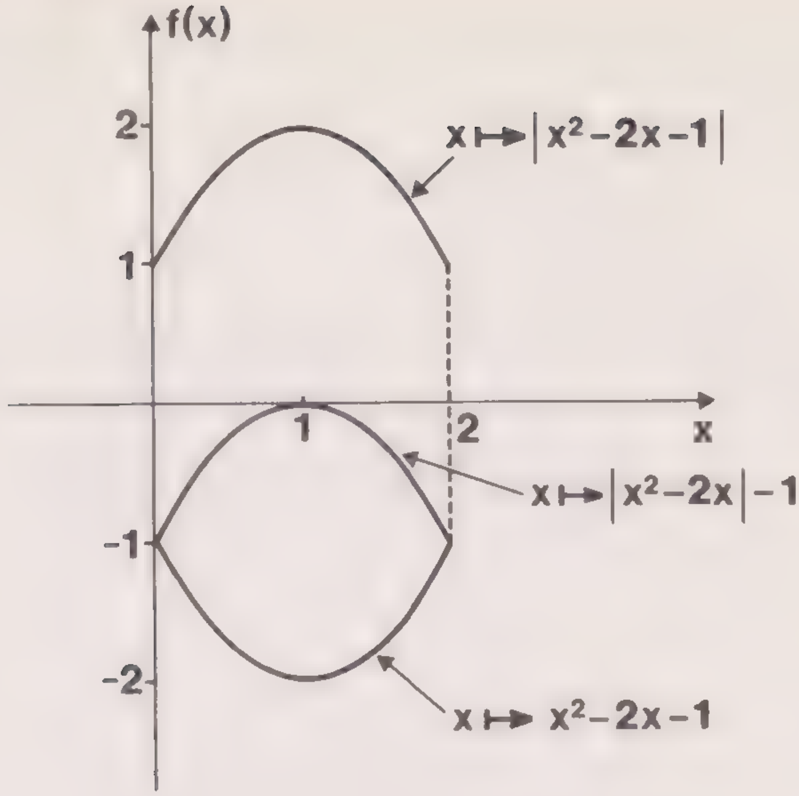
Exercise 1

(i) $-\frac{10}{3}$

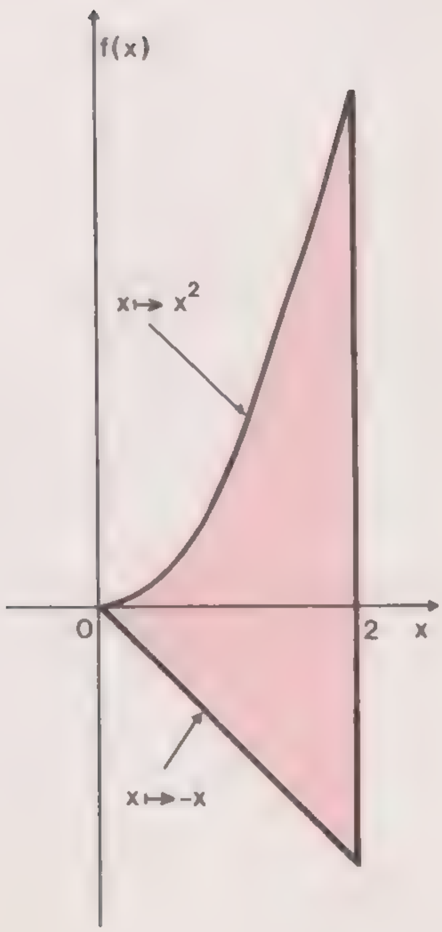
(ii) $\frac{10}{3}$

(iii) $-\frac{2}{3}$

Sketches of the graphs are shown below:



Exercise 2



If the areas above and below the x-axis are calculated separately by definite integrals, we have

$$\int_0^2 (x \longmapsto x^2) + \text{the magnitude of} \left[\int_0^2 (x \longmapsto -x)\right]$$

$$= \tfrac{8}{3} + 2 = \tfrac{14}{3}$$

so the volume is $\frac{14}{3}$ cm^3. We can also represent the whole area by the one definite integral

$$\int_0^2 x \longmapsto (x^2 - (-x)) = \int_0^2 x \longmapsto (x^2 + x) = \tfrac{14}{3},$$

and again we find that the volume is $\frac{14}{3}$ cm^3.

CHAPTER 8 DIFFERENTIATION

8.0 Introduction

In this chapter we shall be concerned with change, and, in particular, the rate at which things change.

Everything in the physical world about us is changing, sometimes rapidly, sometimes very slowly. We know, for example, that plants are growing; so their sizes are changing, albeit very slowly. We know too that an insect's wings move as it flies, even though they usually beat so rapidly that it is impossible for us to follow the movement with our eyes. An intermediate example is the movement of a motor car. Here it is the position of the car that is changing, and we have an instrument in the car which measures the rate of change of position, namely the speedometer.

The general aim of this chapter is to set up a mathematical scheme for describing and measuring rates of change. Because everybody has some familiarity with velocity and acceleration, we have chosen motion as our starting point. The ideas we shall develop are, however, of great generality. They are used not only in kinematics (the study of velocities and accelerations) but also in studying many other types of rate of change, such as the rate of population growth; the rate at which the boiling point of water changes with height on a mountain; the rate at which the temperature in a room decreases with distance from a radiator, and so on. In the first sections we are mainly concerned with the mathematical description of these rates of change, that is, with the concept of the derivative of a function. In later sections we establish some rules for finding derivatives.

As we remarked in Chapter 7, our notation for the calculus is different from the classical notation which is used in most textbooks; our notation is consistent with our basic approach to mathematics through the concept of a function. Once you have mastered the basic principles, there is no objection to your using the traditional Leibniz notation which is discussed in the Appendix.

8.1 Rates of Change

We have already discussed the concept of average velocity in Chapter 4. The purpose of this section is to remind you how average velocity is defined, and to consider some of the consequences of the definition.

To illustrate the idea of average velocity, suppose you made a car journey from London to Edinburgh and recorded the total distance covered at hourly intervals. The recorded data might be given in the following table:

Time (h)	Distance (km)
0	0
1	15
2	110
3	200
4	240
5	300
6	390
7	480
8	570
9	600

Table I

The table shows that at time 0 the distance travelled is 0, where time 0 is the time at which the journey begins. After one hour the car has covered 15 km. After two hours it has covered 110 km, i.e. in the second hour it has covered 95 km. In the third hour it has covered 90 km, but in the fourth and fifth hours it has only covered 40 and 60 km respectively—perhaps you stopped for lunch. Then the car covered 90 km in each of the sixth, seventh and eighth hours of the journey. In the ninth hour it only covered 30 km; perhaps the traffic was heavy near Edinburgh.

In discussing your trip to Edinburgh with a friend who has made a similar long journey you might wish to compare your average velocity with his. To calculate it you would divide the distance travelled by the time:*

$$\text{average velocity} = \frac{\text{distance travelled}}{\text{time}}$$

$$= \frac{600\ \text{km}}{9\ \text{h}}$$

$$= 66.7\ \text{km/h}.$$

Alternatively, you might wish to compare your average velocities over different parts of the journey; for example:

* Velocity means speed in a known direction. Here we assume that the road is straight so that only two directions of motion are possible: away from London and towards it. We distinguish them by giving the velocity a positive sign for motion away from London, and a negative sign for motion towards it.

$$\text{average velocity over first 3 hours} = \frac{\text{distance travelled}}{\text{time}}$$

$$= \frac{200\text{ km}}{3\text{ h}}$$

$$= 66.7\text{ km/h}$$

$$\text{average velocity over last 3 hours} = \frac{600\text{ km} - 390\text{ km}}{9\text{ h} - 6\text{ h}}$$

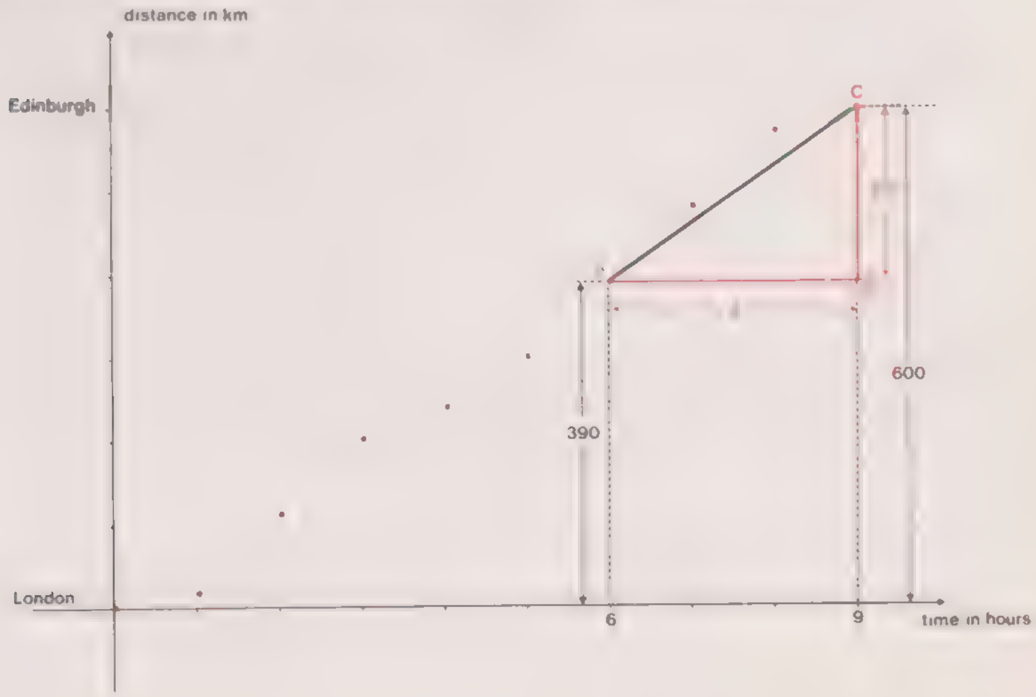
$$= \frac{210}{3}\text{ km h}$$

$$= 70\text{ km/h}.$$

In each case the method of calculation is an application of the formula given in Chapter 4: average velocity over time interval $[t_1, t_2]$ is

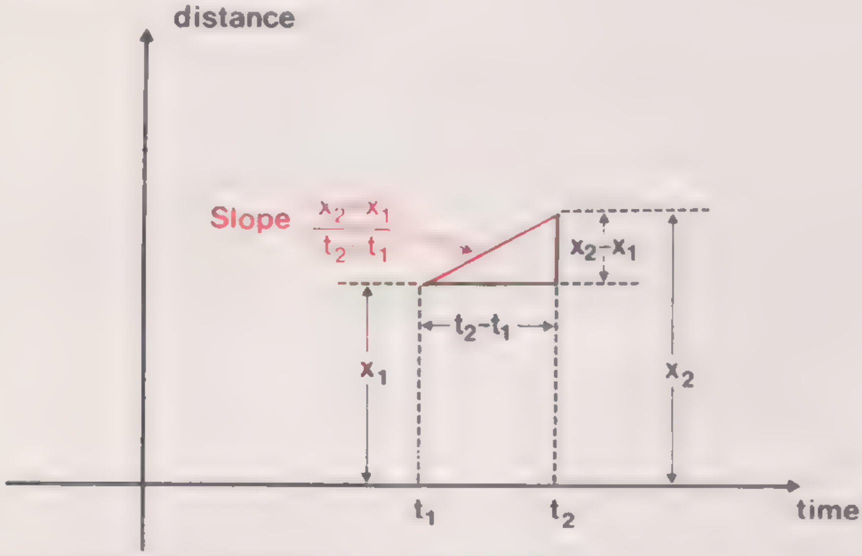
$$\frac{\text{distance travelled in }[t_1, t_2]}{\text{duration of }[t_1, t_2]} = \frac{x_2 - x_1}{t_2 - t_1}$$ Equation (1)

where x_1 and x_2 are the distances from London at the times t_1 and t_2 respectively. The essential restriction on t_1 and t_2 (apart from the obvious one that they must be among the instants for which the distances of the car from London are given) is that they must be different; that is, $t_2 > t_1$, for if $t_2 = t_1$, then the fraction in Equation (1) has denominator zero, and fractions with zero denominator have no meaning.



Average velocities have a very convenient representation in terms of pictorial graphs. The figure above shows how we can use the graph to calculate the average velocity over the time interval [6, 9]. In the right-angled triangle ABC, the side AB corresponds to a time, 3 hours, and BC corresponds to a distance, (600 − 390) km = 210 km. The magnitude of the average velocity over [6, 9] is $\frac{210}{3}$ which is the slope of AC (i.e. tan CAB).

Similarly, the magnitude of the average velocity over any time interval, $[t_1, t_2]$, as given by Equation (1), is the slope of the straight line joining the two points on the pictorial graph corresponding to t_1 and t_2; the unit of the average velocity is km/h.



The concept which underlies average velocity can be generalized to many other situations, if we first pick out its essential feature with the help of the concept of a function. In the example of the car journey, the relevant function is the one tabulated in Table I:

f:(time since start)$\longmapsto$(distance travelled).

In terms of this function, the formula for average velocity is:

$$\text{average velocity over the interval } [t_1, t_2] = \frac{f(t_2) - f(t_1)}{t_2 - t_1}$$ Equation (2)

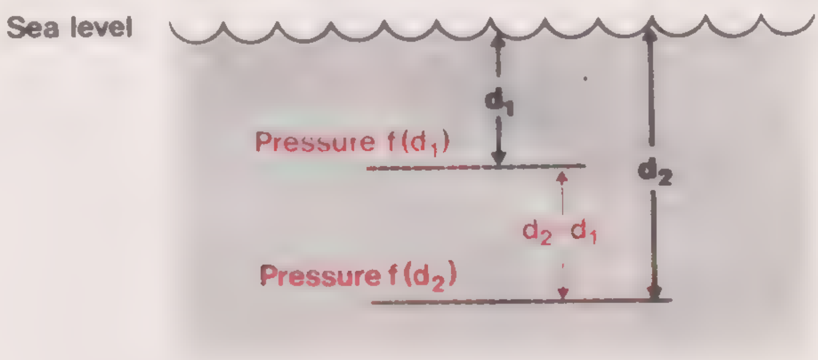
Expressions of the type used in Equation (2) can be useful for many other situations involving rates of change. For example, $f(t)$ may represent the quantity of water in a reservoir at time t. In this case, the expression on the right of Equation (2) has the interpretation

$$\frac{f(t_2) - f(t_1)}{t_2 - t_1} = \frac{\text{change in quantity of water in reservoir during } [t_1, t_2]}{\text{duration of } [t_1, t_2]}.$$

That is, it gives the average rate at which the quantity of water in the reservoir changes during $[t_1, t_2]$. As another example, f might be the function:

$$f\colon(\text{depth below sea level}) \longmapsto (\text{hydrostatic pressure})$$

(at some particular time and place). In this case, if d_1 and d_2 are two depths,



then

$$\frac{f(d_2) - f(d_1)}{d_2 - d_1} = \frac{\text{corresponding change in pressure}}{\text{change in depth}}$$

$$= \text{average rate of change of pressure with depth.}$$

The value of this fraction is roughly 0.1 atmosphere per metre. An aquanaut can use this value to deduce his depth changes from measured changes in pressure.

In general, if f is any real function (i.e. a function whose domain and codomain are subsets of R), we define:

Average rate of change of $f(t)$ over the interval $[t_1, t_2]$ is

$$\frac{f(t_2) - f(t_1)}{t_2 - t_1}.$$

In the car example, the distance travelled between times t and $t + h$ is $f(t + h) - f(t)$. Therefore the average velocity in this time interval is

$$\frac{f(t + h) - f(t)}{h} \quad (h \in R^+)$$

If we were interested in the average velocity in the interval $[t - h, t]$ we would have the expression

$$\frac{f(t) - f(t - h)}{h} \quad (h \in R^+)$$

To avoid having two different expressions we adopt the former and allow h to range over all the real numbers except zero. (To see that this is justified, put $h = -k$ in the first expression to give

$$\frac{f(t-k)-f(t)}{-k}$$

$$=\frac{f(t)-f(t-k)}{k}$$

which has the same form as the second expression.)
Suppose now that h is very small. The nearer h is to zero, the better idea we shall have of the average velocity of the car near to the time t. If the expression

$$\frac{f(t+h)-f(t)}{h}$$

tends to a limit L as h tends to zero, then we say that L is the velocity of the car *at time t*.

This velocity is often called the *instantaneous velocity*. We write

$$v(t) = \lim_{h \to 0} \frac{f(t+h)-f(t)}{h}$$

Exercise 1

Find the instantaneous velocity at time t of a car which moves in such a way that its distance from some fixed point on the road it travels is at^3 where a is some positive number.

8.2 The Derivative

Just as the idea of average velocity can be generalized to give a definition of average rate of change for any function, so the idea of an instantaneous velocity can also be generalized to give a definition of rate of change of a real function having no direct connection with kinematics. As in the discussion of average rates of change, all that we have to do is to apply the same formula as in the definition of instantaneous velocity, and call the analogue of instantaneous velocity the (instantaneous) rate of change. That is to say, if f is any real function, we can define the (instantaneous) rate of change of $f(x)$ at the element x in the domain of f, to be the number

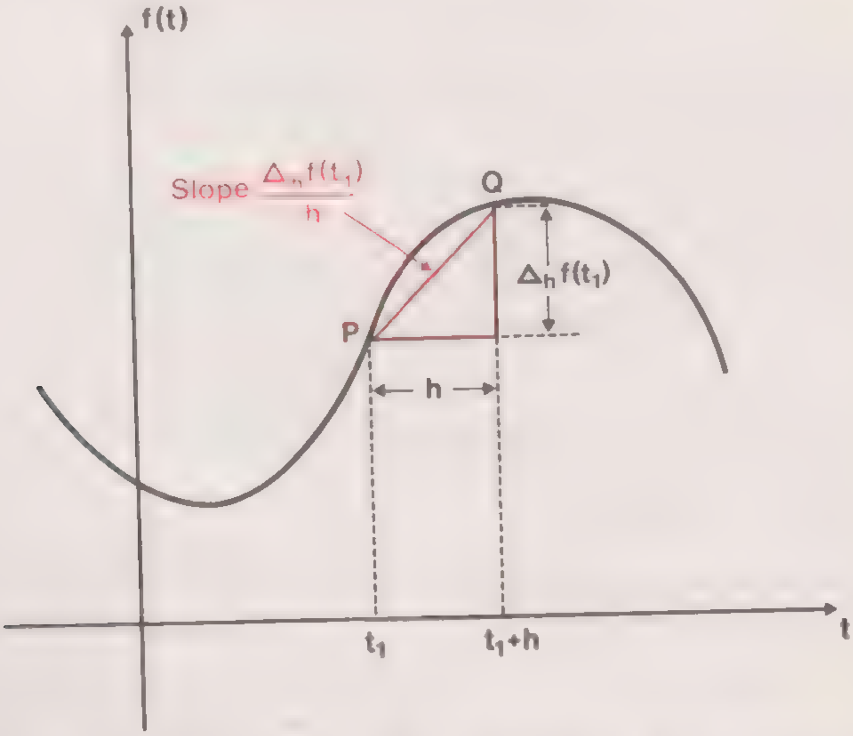
$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Definition 1

provided that this limit exists. This rate of change at x is usually called the derivative of f at x and it is denoted by $f'(x)$. Thus the velocity of a car t seconds after the start of its journey is equal to the derivative at t of the function (numbers of seconds since starting) $\longmapsto$ (distance since starting).

There are other notations for the derivative: in the most important of them, the Leibniz notation, we write $\dfrac{df(x)}{dx}$ in place of $f'(x)$.

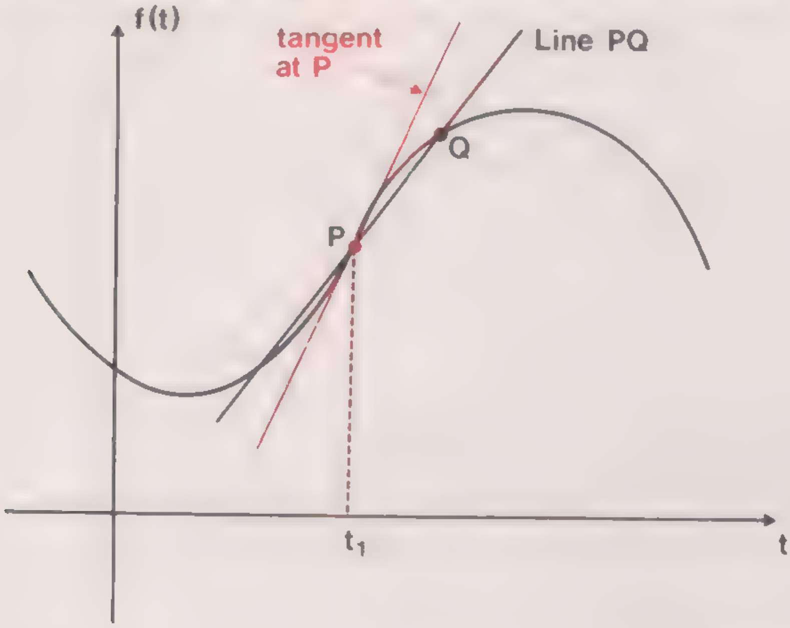
Like the average rate of change, the instantaneous rate of change, or derivative, has a very useful interpretation in terms of a pictorial graph of the function. We have already seen that the average rate of change of $f(t)$ over an interval $[t_1, t_1 + h]$ equals the slope of the straight line joining the points on the graph corresponding to t_1 and $t_1 + h$.



(Notice that in this diagram we have introduced the symbols $\Delta_h f(t_1)$ to denote the difference $f(t_1 + h) - f(t_1)$. We shall continue to use the symbol Δ_h as appropriate.)

To obtain the derivative at t_1, we let the magnitude of h get smaller and smaller. In the above figure the point denoted by Q then slides (like a bead) along the curve towards the point P which remains fixed. As Q approaches P, the line rotates about P, and approaches a limiting position which we define to be the tangent to the curve at P. It follows that when h is very small, the slope of the line PQ is very close to the

slope of the tangent at P, and consequently that the limit of the slope of the line PQ near P is equal to the slope of the tangent at P.



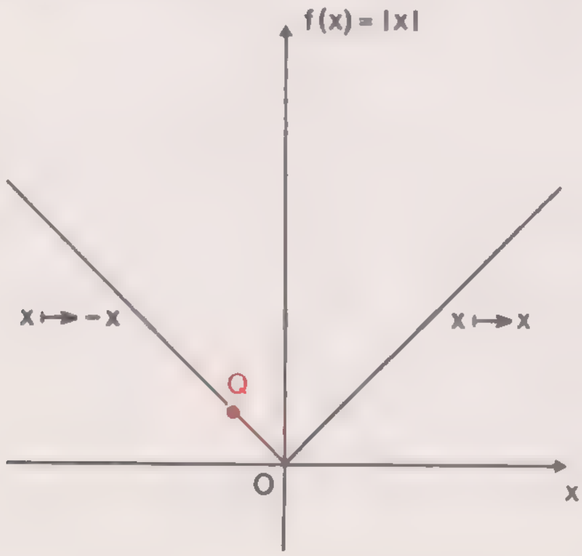
In symbols, using Definition 1, we have: (slope of tangent to graph at P) = (derivative of function at t_1)

The geometrical interpretation of the derivative is very useful, not only in geometry. We shall use it frequently when discussing applications of the derivative; for example, when obtaining approximations to real functions and when finding the maximum and minimum values of the images of real functions.

There are functions whose graphs do not have tangents at every point. For example, the graph of the function

$$f: x \longmapsto |x| \quad (x \in R)$$

does not have a tangent at (0, 0).



You can see this by considering the slope of the line OQ as Q approaches the origin, O, where Q is a point on the graph. Consider first the case in which Q is a point on the graph of the function $x \longmapsto -x$, and then that when Q is a point on the graph of the function $x \longmapsto x$. The slopes of the lines are respectively -1 and $+1$; the tangent at $(0, 0)$ does not exist. We have

$$\lim_{h \to 0} \frac{\Delta_h f(x)}{h} = \begin{matrix} +1 \text{ if } x > 0 \\ -1 \text{ if } x < 0 \end{matrix} \text{ but } \lim_{h \to 0} \frac{\Delta_h f(0)}{h} \text{ does not exist.}$$

We say that the derivative of f at 0 does not exist.

Exercise 1

Which of the following statements are true?

(i) The tangent to a curve at P cannot cross the curve at P.
(ii) The angle between the tangent to a curve at P and the horizontal axis is the limit, as Q approaches P along the curve, of the angle between the line PQ and this axis.
(iii) The tangent to a curve at P may be defined as the straight line that meets the curve only at P.
(iv) If f is continuous at every element in its domain there is a tangent at every point on its graph.

Exercise 2

Show that the derivative of a constant function at any element in its domain is 0.
A constant function is one for which the image of every element of the domain is the same.

Exercise 3

Sketch the graph of $f: t \longmapsto t^2 \quad (t \in R)$.
Find $f'(t)$ and evaluate $f'(-3)$, $f'(0)$ and $f'(2)$.

What can you deduce about the slopes of the tangents to the graph of f at -3, 0 and 2?

The Derived Function

We have already indicated the wide range of applications of the concept of derivative which we defined in the preceding section. To make full use of

this concept we need to look at some of the properties of the derivative. The first fruits of this study will be a set of rules which will enable us to calculate derivatives of functions quickly and easily, without having to go back to the definition every time.

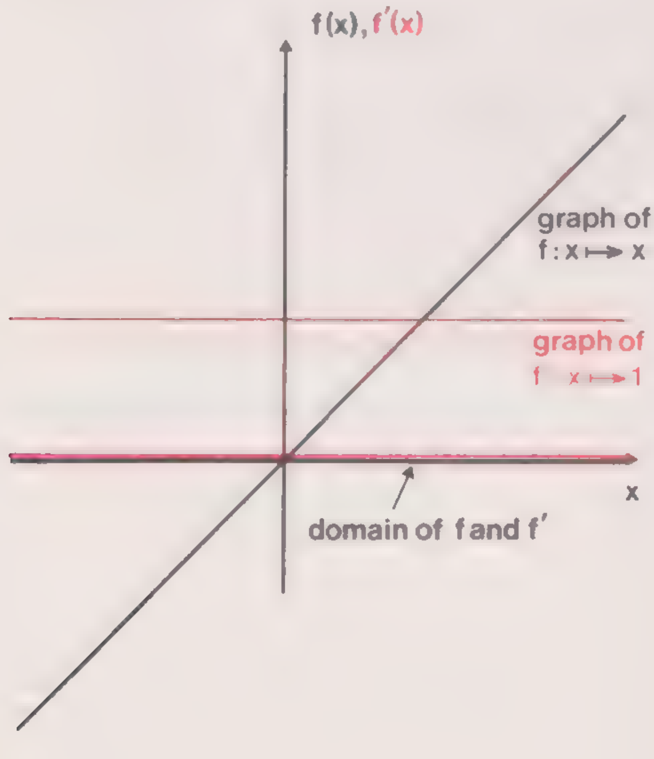
We start with Definition 1. The derivative of a real function, f, at x:

$$f'(x) = \lim_{h \to 0} \frac{\Delta_h f(x)}{h} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.$$

The derivative is a number; however, since this number depends on the value of x, we can use the notion of derivative to define a function. This function maps x to the value of the derivative of f at x. In fact, by our notation $f'(x)$ we have already implicitly recognized the existence of this function. This new function f' is called the derived function of f, and in finding it we are said to differentiate the function f. Definition 2

The domain of the derived function will be taken to comprise all the values of x for which $f'(x)$ exists. Since Definition 2 involves f, the domain of the derived function must be a subset of the domain of f. In some cases the two domains may be the same; in others, the domain of f' may be a proper subset of that of f; that is, it may exclude certain numbers that are in the domain of f, because the limit defining the derivative does not exist at these numbers. We say that a function f is differentiable at those elements in its domain where $f'(x)$ exists. Thus the domain of f' is that subset of the domain of f comprising the numbers at which f is differentiable.

Example 1



The derived function f' of the function

$$f : x \longmapsto x \qquad (x \in R)$$

is given by

$$f'(x) = \lim_{h \to 0} \frac{(x + h) - x}{h} = \lim_{h \to 0} \frac{h}{h} = 1$$

and the limit exists for all $x \in R$: so the domain of f' is also R.

Example 2

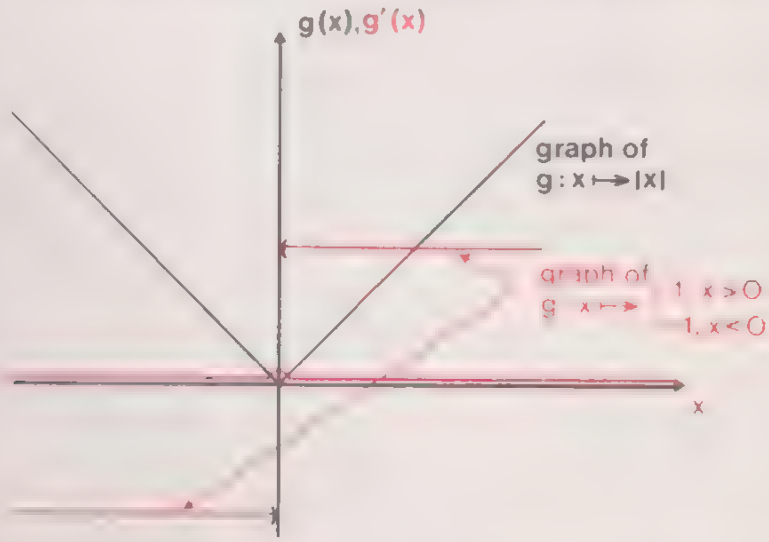
The derived function g' of the modulus function

$$g : x \longmapsto |x| \qquad (x \in R)$$

is given (where it exists) by:

$$g'(x) = \lim_{h \to 0} \frac{|x + h| - |x|}{h}.$$

The evaluation of this limit, though not difficult, is a little tedious (see text following this example). A simpler method of finding the domain of g' is to use the graph of g. The graph below shows the point $(0, 0)$ at which there is no tangent; since the slope of the tangent coincides with the value of the derivative, we see that there is no derivative, i.e. the limit does not exist near $x = 0$. At all other points there is a tangent (in fact it coincides with part of the graph) and so the function is differentiable everywhere except at $x = 0$. Thus the domain of g' consists of the set R with 0 omitted.



Like all arguments based on graphs (or figures in general), the one given in Example 2 makes use of geometrical intuition and it is therefore a demonstration rather than a proof. For a proof we must go back to the definition and evaluate the limit. We consider three cases:

(i) $x > 0$

In evaluating the limit it is only small values of $|h|$ that matter: so we need only consider $h > -x$, so that $h + x > 0$. It follows that $|x| = x$ and $|x + h| = x + h$; the definition of $f'(x)$ therefore gives

$$f'(x) = \lim_{h \to 0} \frac{(x + h) - x}{h} = \frac{h}{h} = 1.$$

(ii) $x < 0$

Again, we need only consider small values of $|h|$. We take $h < -x$, so that $h + x < 0$. It follows that $|x| = -x$ and $|x + h| = -(x + h)$; so the expression for $f'(x)$ is

$$f'(x) = \lim_{h \to 0} \frac{|x + h| - |x|}{h} = \lim_{h \to 0} \frac{-(x + h) + x}{h} = \frac{-h}{h} = -1.$$

(iii) $x = 0$

The expression for $f'(x)$ is

$$f'(x) = \lim_{h \to 0} \frac{|h|}{h},$$

but since $\frac{|h|}{h}$ takes the value $+1$ for small positive h and -1 for small negative h, there is no number close to $\frac{|h|}{h}$ for *all* small h: that is, *the function* $x \longmapsto |x|$ *is not differentiable at* 0.

8.3 Differentiation of Polynomials

You have already seen how to differentiate a few functions. The results are shown in the table: in each case f has domain R.

f	f'	
$x \longmapsto \text{constant}$	$x \longmapsto 0$	$(x \in R)$
$x \longmapsto x$	$x \longmapsto 1$	$(x \in R)$
$x \longmapsto x^2$	$x \longmapsto 2x$	$(x \in R)$
$x \longmapsto x^3$	$x \longmapsto 3x^2$	$(x \in R)$
$x \longmapsto \|x\|$	$\begin{cases} x \longmapsto 1 & (x \in R^+) \\ x \longmapsto -1 & (x \in R^-) \end{cases}$	

Exercise 1

Differentiate the function:

$$f: x \longmapsto ax^2 + bx + c \quad (x \in R).$$

To work out each new derivative, when it is required, from the definition

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{\Delta_h f(x)}{h}$$

would be very laborious. It is much easier to work out in advance a system of rules which makes it possible to differentiate many functions from a knowledge of relatively few basic derivatives. To formulate these rules, we begin with a class of particularly simple functions, namely the *polynomial functions.*

A polynomial function is any function of the form·

$$p: x \longmapsto a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \qquad (x \in R)$$

where $a_0, a_1, \ldots, a_n \in R$.

Rather than try to differentiate the general polynomial function of degree n immediately, let us begin with the simplest such polynomial function which is

$$x \longmapsto x^n \quad (x \in R)$$

where n is any positive integer or zero. The cases $n = 0, 1, 2, 3$, have already been treated in various exercises and examples; so if you know the binomial theorem you should not find the general case difficult.

Exercise 2

Differentiate the function:

$$f: x \longmapsto x^n \quad (x \in R)$$

where n is any positive integer or zero.

Extending the f' notation for derived functions, we can write the result of Exercise 2:

$$(x \longmapsto x^n)' = (x \longmapsto nx^{n-1}) \quad (x \in R).$$

Some of the results obtained earlier (for example, the derivative of a constant function is zero, the derivative of $t \longmapsto t^3$ is $t \longmapsto 3t^2$) are

special cases of this important result. The general polynomial

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

can be built up from "elementary" polynomials of the form x^k by first multiplying each of these elementary polynomials by the appropriate coefficient a_k to get one of the terms $a_k x^k$ in the general polynomial, and then adding all these terms together. By the corresponding operations on the "elementary" polynomial functions $x \longmapsto x^k$ we can build up the general polynomial function:

$$p : x \longmapsto a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \quad (x \in R).$$

It follows that we shall also be able to build up the derived function p' provided we can obtain rules enabling us to deduce:

(i) the derived function of $x \longmapsto a_k x^k$ from that of $x \longmapsto x^k$,
(ii) the derived function of a sum of functions from the derived functions of its individual terms.

Exercise 3

Using the definition of a derivative show that

(i) the derived function of $t \longmapsto af(t)$ is $t \longmapsto af'(t)$ where $t \in$ (domain of f');

(ii) for any two functions f and g with the same domain, the derived function of $t \longmapsto f(t) + g(t)$ is $t \longmapsto f'(t) + g'(t)$ where $t \in$ (domain which is common to both f' and g').

We can state the two rules of differentiation given in Exercise 3 as follows:

First Rule of Differentiation

$(af)' = af'$, i.e. multiplying any function by a number multiplies its derived function by the same number. **Rule 1**

Second Rule of Differentiation

$(f + g)' = f' + g'$, i.e. the derived function of a sum of functions is the sum of the individual derived functions, provided the domains are appropriate. **Rule 2**

The second rule can be extended to cover any (finite) number of terms in a sum. We can now construct a derivative for the general polynomial

$$p : x \longmapsto a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \quad (x \in R).$$

We have already shown that

$$(x \longmapsto x^k)' = x \longmapsto kx^{k-1} \quad (x \in R).$$

By Rule 1, multiplying a function by a constant (in this case a_k), multiplies its derived function by the same constant, so

$$(x \longmapsto a_k x^k)' = x \longmapsto k a_k x^{k-1} \quad (x \in R).$$

By Rule 2, the derived function of a sum of two or more functions is the sum of their individual derived functions. The polynomial function p is the sum of the functions $(x \longmapsto a_n x^n)$, $(x \longmapsto a_{n-1} x^{n-1})$ etc.; therefore its derivative is the sum of their derivatives:

$$p' = (x \longmapsto n a_n x^{n-1}) + (x \longmapsto (n-1) a_{n-1} x^{n-2}) + \cdots$$
$$+ (x \longmapsto a_1)$$

that is,

$$p' = x \longmapsto (n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \cdots + 2a_2 x + a_1) \quad (x \in R).$$

Exercise 4

(i) Differentiate

$$x \longmapsto 10x^5 + \tfrac{1}{8}x^3 + x \quad (x \in R).$$

(ii) Differentiate $x \longmapsto 2x + |x| \quad (x \in R)$

8.4 The Differentiation Operator

We have seen in section 8.2 that to every function f there corresponds a unique derived function, f', whose domain is a subset of the domain of f. (In some cases, this subset contains no elements, that is to say f cannot be differentiated anywhere. Can you think of an example? In case you cannot, you will find one in Example 1 below.)

That is, we have a rule which assigns to each member, f, of the set of real functions a unique function f'; the rule can be described by the word *differentiate*, and you may find it helpful to think of it as a machine which takes one function and turns it into another, as illustrated below:

INPUT MACHINE OUTPUT

$$\xrightarrow[f]{} \boxed{\text{DIFFERENTIATE}} \xrightarrow[f']{}$$

Example 1

An example of a function which cannot be differentiated anywhere in its

domain is a function whose domain consists of isolated numbers, for example:

$$f: x \longmapsto x \quad (x \in \{1, 2, 3, 4\}).$$

The reason why this function cannot be differentiated is that the definition of a derivative at x involves taking the limit of

$$\frac{f(x+h) - f(x)}{h}$$

as h tends to zero. But this means that there must be some interval $[x - h, x + h]$ contained in the domain of f; otherwise we cannot apply our definition of a limit. But if x is one of the numbers 1, 2, 3 or 4, and $0 < |h| < 1$, then $x + h$ cannot be one of these numbers, so $f(x + h)$ is not defined; and hence the derivative at x does not exist.

Earlier, we have introduced

$$\Delta_h f(x)$$

to mean

$$f(x+h) - f(x).$$

We call Δ_h the *difference operator*,* and we can write:

$$\Delta_h : f \longmapsto [x \longmapsto f(x+h) - f(x)] \quad (f \in F)$$

where F is the set of all real functions.

Here we shall do a similar thing: we define an operator D whose effect on any function in its domain is to differentiate it:

$$D : f \longmapsto f'.$$

It is called the differentiation operator. To complete the definition of the operator D we must specify its domain. We shall take this domain, as for Δ_h, to be the set of all real functions, F. Using the operator D we can write results obtained earlier as follows:

$$D(x \longmapsto x^m) = x \longmapsto mx^{m-1}$$

$$D(af) = aDf$$

$$D(f + g) = Df + Dg$$

* The word *operator* is used for a mapping whose domain is itself a set of functions.

Exercise 1

It is usual to write D^2 for $D \circ D$, D^3 for $D \circ D \circ D$, and so on.

(i) Evaluate $D^2\,(x \longmapsto ax^2 + bx + c)$
(ii) Evaluate $D^3\,(x \longmapsto x^3)$.

If f is any real function, the function $D^2 f$ is called the second derived function of f, $D^3 f$ is called the third derived function, and so on.

In the notation used in section 8.2, we would write: f'' for $D^2 f$; f''' or $f^{(3)}$ for $D^3 f$; $f^{(4)}$ for $D^4 f$; $f^{(n)}$ for $D^n f$, and so on.

8.5 Differentiation of Products

Differentiating the product of two functions is not quite as easy as differentiating their sum.

Exercise 1

Show by an example (as simple as possible) that cases exist where

$$D(f \times g) \neq Df \times Dg$$

We have already seen that

$$D(af) = aDf.$$

As long as we want to differentiate polynomial functions only, this is all we need to know about the derived function of a product, but to differentiate a function such as

$$x \longmapsto x \sin x \quad (x \in R)$$

which is a product of the two functions

$$x \longmapsto x \qquad (x \in R)$$

and $\quad x \longmapsto \sin x \quad (x \in R)$

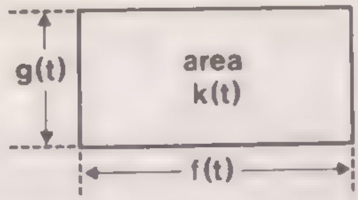
neither of which is a constant, we need a more general rule for differentiating products of functions.

Let us denote the two functions whose product we wish to differentiate by f and g, and their product by k, so that

$$f \times g = k.$$

We assume for simplicity that the functions have domain R. To illustrate the product relationship and the subsequent argument, we suppose that

the elements of the domain of f, g and k are times (represented by a variable t), and that the images of t under the functions f, g and k are the sides and area of a rectangle:



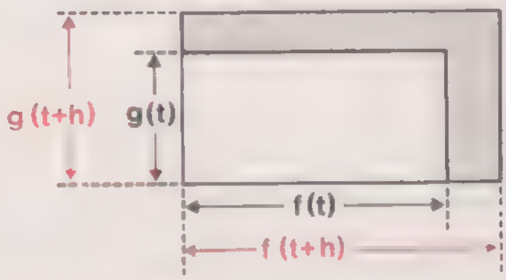
Our problem is to evaluate $k'(t)$, the rate at which the area changes with time. By the definition of a derivative, this is

$$k'(t) = \lim_{h \to 0} \frac{k(t + h) - k(t)}{h}$$

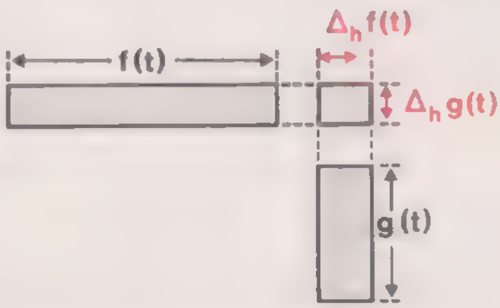
that is

$$k'(t) = \lim_{h \to 0} \frac{f(t + h)g(t + h) - f(t)g(t)}{h}.$$ Equation (1)

The numerator is the difference in area of two rectangles:



and is therefore equal to the area of the shaded L-shaped strip in the above diagram. This strip can be treated as the sum of three rectangles, as shown below:



(Remember that $\Delta_h f(t) = f(t + h) - f(t)$.)

Adding the areas of the rectangles and substituting in Equation (1), we get

$$k'(t) = \lim_{h \to 0} \frac{f(t)\Delta_h g(t) + g(t)\Delta_h f(t) + \Delta_h g(t)\Delta_h f(t)}{h}$$

$$= f(t)\lim_{h\to 0}\frac{\Delta_h g(t)}{h} + g(t)\lim_{h\to 0}\frac{\Delta_h f(t)}{h} + \lim_{h\to 0}\frac{\Delta_h g(t)\Delta_h f(t)}{h},$$

$$= f(t)g'(t) + g(t)f'(t) + \lim_{h\to 0}\frac{\Delta_h g(t)\Delta_h f(t)}{h},$$

by the definition of a derivative.

The last term on the right-hand side can be written

$$\lim_{h\to 0}\frac{\Delta_h f(t)}{h} \times \lim_{h\to 0}\Delta_h g(t) = f'(t) \times 0 = 0,$$

since both these limits exist and

$$\lim_{h\to 0}\Delta_h g(t) = \lim_{h\to 0}[g(t+h) - g(t)] = 0,$$

if g is a continuous function.* (In fact, this is not a new assumption since we have already assumed that g is differentiable, and it can be proved without much difficulty that every differentiable function is continuous: a result which is intuitively obvious if we regard differentiability in terms of being able to draw tangents.)

We therefore have

$$k'(t) = f(t)g'(t) + g(t)f'(t).$$ **Equation (2)**

That is, the rate of change of the area of the rectangle is the sum of two terms: the rate at which area is being added because the length of the side of original length $f(t)$ is increasing, plus the rate at which area is being added because the length of the side of original length $g(t)$ is increasing.

Equation (2) is the rule for differentiating products. We have thought of t as a time in deriving this rule, but the rectangle illustration can be dropped and the mathematical argument applies to any product of functions with domain R or some subset of R. The Product Rule can, of course, also be written in terms of the functions themselves instead of images; it then takes the form:

$$(fg)' = fg' + gf'$$

or, in terms of D:

$$D(fg) = fDg + gDf.$$

This can be regarded as a generalization of the equation $D(af) = aDf$, to which it reduces when g is a constant function.

* Notice that we have to assume that g is continuous in order to be sure that $\lim_{h\to 0} g(t+h) = g(t)$. If g is not continuous, then this limit may exist and be different from $g(t)$

Exercise 2

Verify the product rule by considering the product of the two functions:

$$f: x \longmapsto 2x + 1 \qquad (x \in R)$$

$$g: x \longmapsto x - 1 \qquad (x \in R).$$

8.6 Differentiation of Composite Functions

We have seen that for addition of functions

$$D(f + g) = Df + Dg$$

and that for multiplication of functions

$$D(f \times g) \neq Df \times Dg$$

We shall consider division later. Subtraction follows directly from addition to give

$$\begin{aligned} D(f - g) &= D(f + (-g)) \quad \text{where } -g: x \longmapsto -g(x) \\ &= Df + D(-g) \\ &= Df - Dg \end{aligned}$$

using the rule $D(ag) = aDg$, with $a = -1$.

The last of the binary operations on functions that we defined in Chapter 3, was the operation of composition

$$f \circ g: x \longmapsto f(g(x)).$$

We approach the investigation of composition in the same way as we tackled that of multiplication.

Exercise 1

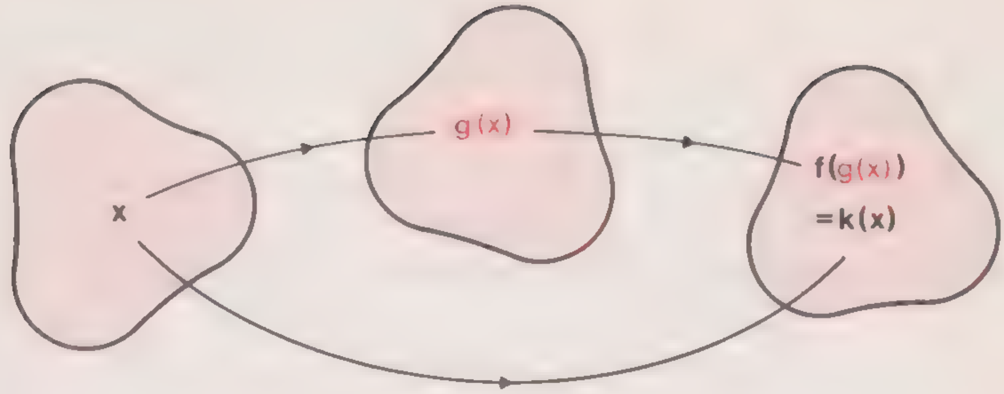
Show by as simple an example as possible, that cases exist where

$$D(f \circ g) \neq Df \circ Dg.$$

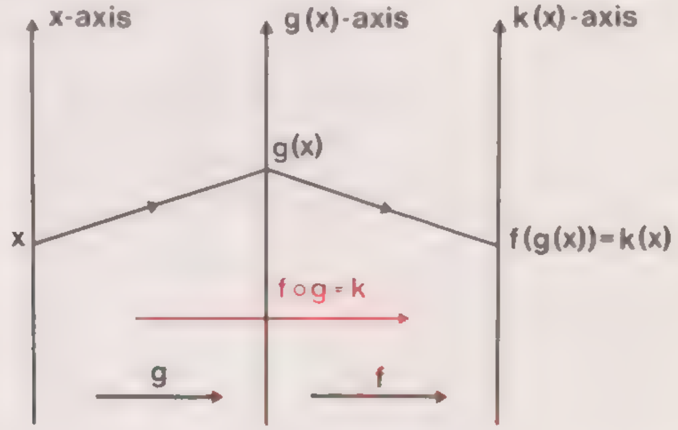
So we must try to find a rule for differentiating the composite of two functions f and g. We write

$$f \circ g = k, \quad \text{i.e. } f \circ g(x) = f(g(x)) = k(x).$$

The relation between these functions is illustrated below:



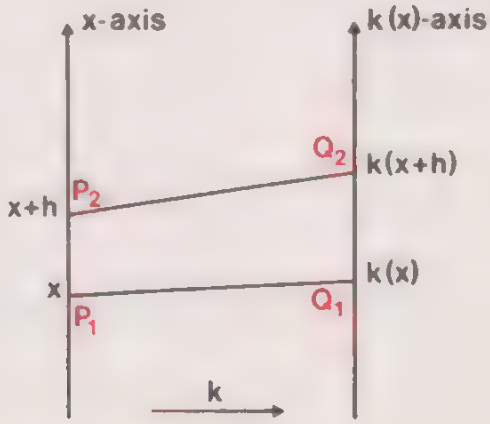
It can also be illustrated in this way:



To determine whether $f \circ g$ is differentiable we consider the formula defining a derivative:

$$k'(x) = \lim_{h \to 0} \frac{k(x+h) - k(x)}{h}.$$

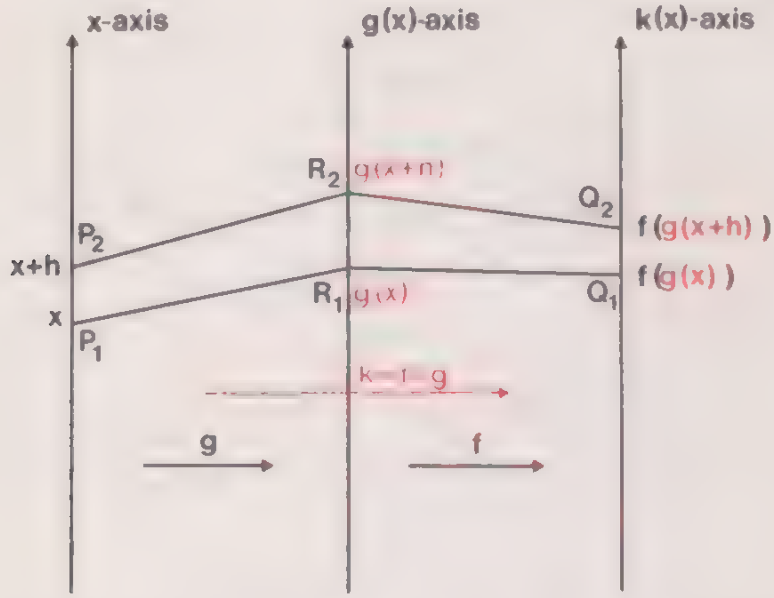
The limit on the right can be interpreted using a diagram similar to the one below:



We see that $\dfrac{k(x+h) - k(x)}{h} = \dfrac{Q_1Q_2}{P_1P_2}$.

We shall call this the magnification produced by the function k when it maps the segment P_1P_2 (i.e. the interval $[x, x+h]$) to the segment Q_1Q_2 (i.e. the interval $[k(x), k(x+h)]$).

Our problem is to express the derived function of $f \circ g$ in terms of the functions f and g: in terms of diagrams, it is to combine the last two diagrams. We can combine the two diagrams in the following way:



Instead of performing the mapping k in one stage we perform it in two stages: we perform first g and then f. Thus the segment P_1P_2 is mapped first to the segment R_1R_2 and then to the segment Q_1Q_2. The magnifications produced in these two steps are

$$\frac{R_1R_2}{P_1P_2} \quad \text{and} \quad \frac{Q_1Q_2}{R_1R_2}$$

respectively, and the overall magnification is the product of these two:

$$\frac{Q_1Q_2}{P_1P_2} = \frac{R_1R_2}{P_1P_2} \times \frac{Q_1Q_2}{R_1R_2}. \qquad \text{Equation (1)}$$

We have

$$\lim_{h \to 0} \frac{Q_1Q_2}{P_1P_2} = k'(x).$$

For the magnification in the first stage of the two-stage process we have

$$\lim_{h \to 0} \frac{R_1R_2}{P_1P_2} = g'(x).$$

* This step really covers up the difficult part of the verification, which depends on the continuity of g.

Finally, the magnification produced by the function f has the limit

$$\lim_{h \to 0} \frac{Q_1 Q_2}{R_1 R_2} = f'(g(x))^*.$$

This time the derivative has to be computed at $g(x)$ and not at x, since the point R_1 corresponds to the number $g(x)$.

If we put these three limits together we obtain from Equation (1) the Composite Function Rule:

$$k'(x) = g'(x) \times f'(g(x)).$$

In function notation, we have:

$$(f \circ g)' = (f' \circ g) \times g'.$$

It is often called the chain rule, or "function of a function" rule, since a composite function is, in a sense, a function of a function.

Example 1

Differentiate

$$k: x \longmapsto (2x + 1)^2 \quad (x \in R).$$

(This could be done by expanding the brackets, but we shall apply the above rule.)

If we let $g: x \longmapsto 2x + 1 \quad (x \in R)$
and $f: x \longmapsto x^2 \quad (x \in R)$
then $k = f \circ g$.

We can now use the composite function rule, i.e.

$$\begin{aligned} k' &= (f' \circ g) \times g' \\ &= [(x \longmapsto 2x) \circ (x \longmapsto 2x + 1)] \times (x \longmapsto 2) \\ &= x \longmapsto 2(2x + 1) \times 2 \\ &= x \longmapsto 4(2x + 1) \qquad (x \in R). \end{aligned}$$

Exercise 2

If $f: x \longmapsto x^2 + 1 \quad (x \in R)$
and $g: x \longmapsto x^2 - 1 \quad (x \in R)$,
find $f \circ g$.

Differentiate $f \circ g$ directly and also using the composite function rule.

8.7 Differentiation of Quotients

In this section and the next we complete the list of rules of differentiation by deducing from the product rule a rule for differentiating the quotient of two functions, and deducing from the composite function rule a rule for differentiating inverse functions. We begin by finding the derived function of the reciprocal function, which we denote here by the letter r, so that

$$r : x \longmapsto \frac{1}{x} \qquad (x \in R, \quad x \neq 0);$$

then

$$\begin{aligned} r'(x) &= \lim_{h \to 0} \left(\frac{r(x+h) - r(x)}{h} \right) \\ &= \lim_{h \to 0} \frac{1}{h} \left(\frac{1}{x+h} - \frac{1}{x} \right) \qquad \text{(by definition of } r\text{)} \\ &= \lim_{h \to 0} \frac{1}{h} \left(\frac{x - (x+h)}{(x+h)x} \right) \\ &= \lim_{h \to 0} \left(\frac{-h}{h(x+h)x} \right) \\ &= \lim_{h \to 0} \left(\frac{-1}{(x+h)x} \right) \qquad \text{(the cancellation of the } h\text{'s is valid since } h \neq 0\text{).} \end{aligned}$$

Taking the limit as $h \rightsquigarrow 0$ we obtain (since $x \neq 0$)

$$r'(x) = \frac{-1}{x^2}$$

i.e.

$$D\left(x \longmapsto \frac{1}{x} \right) = x \longmapsto \frac{-1}{x^2}$$

or

$$r' = -r \times r,$$

the domain being the set of non-zero real numbers throughout.

Note

(i) We shall denote the product function $r \times r$ by r^2.
Thus

$$r^2 : x \longmapsto r(x) \times r(x) = (r(x))^2.$$

(ii) If

$$v : x \longmapsto v(x), \qquad (x \in R)$$

then we shall denote the function:

$$x \longmapsto \frac{1}{v(x)} \qquad (x \in R, \quad v(x) \neq 0)$$

by $\frac{1}{v}$.

We can now apply our last result to the differentiation of the quotient of two functions. If two functions u and v have as domain the same subset of R, and codomain R, then we define their quotient, $\frac{u}{v}$ or u/v, by

$$\frac{u}{v} : x \longmapsto \frac{u(x)}{v(x)} \quad (x \in \text{domain of } u \text{ and } v,\ v(x) \neq 0).$$

The formula for differentiating such functions is a useful one. It may help you to remember the rule if you work it out for yourself: so the rest of the derivation of this formula is set as an exercise.

Exercise 1

Denoting the reciprocal function by r as in the text, we can write

$$\frac{u}{v} = u \times \frac{1}{v} = u \times (r \circ v).$$

Using the following results:

(i) the derived function of r is $-r^2$.

(ii) the product rule, i.e.

$$(f \times g)' = f' \times g + f \times g',$$

(iii) the composite function rule, i.e.

$$(f \circ g)' = (f' \circ g) \times g',$$

express $\left(\frac{u}{v}\right)'$ in terms of u', v', u and v.

The result of the last exercise gives the Quotient Rule:

$$\left(\frac{u}{v}\right)' = \frac{u' \times v - u \times v'}{v^2}$$

which means that

if $$w(x) = \frac{u(x)}{v(x)}$$

then $$w'(x) = \frac{u'(x) \times v(x) - u(x) \times v'(x)}{(v(x))^2}.$$

Exercise 2

Differentiate:

$$x \longmapsto \frac{1}{x^k} \qquad (x \in R, \quad x \neq 0)$$

where k is a positive integer.

Exercise 3

Differentiate:

(i) $w_1 : x \longmapsto \dfrac{-2}{2x+1} \qquad (x \in R, \quad x \neq -\tfrac{1}{2}).$

(ii) $w_2 : x \longmapsto \dfrac{2x-1}{2x+1} \qquad (x \in R, \quad x \neq -\tfrac{1}{2}).$

Can you see why the two results are the same?

8.8 Differentiation of Inverse Functions

The last rule of differentiation we shall consider is the rule for differentiating the inverse of a function. The inverse of a one–one function, as defined in Chapter 3, is a function that reverses the effect of the original one. For example, the inverse of the function "double the number" is the function "halve the number", since if we double a number and halve the result we come back to the original number. We saw in Chapter 3, that a function has an inverse if and only if it is one–one; that is, if each element in the domain has a different image.

It follows that if g is the inverse of f, then

$$g \circ f : x \longmapsto x \qquad (x \in \text{domain of } f)$$

and $$f \circ g : x \longmapsto x \qquad (x \in \text{domain of } g).$$

We can use the composite rule to find the derivative of g in terms of

$$f \circ g = x \longmapsto x$$

therefore $\qquad (f \circ g)' = x \longmapsto 1$

and by the composite rule, $(f \circ g)' = (f' \circ g) \times g'$

which gives us the Inverse Function Rule:

$$g' = \frac{x \longmapsto 1}{f' \circ g} = \frac{1}{f' \circ g}.$$

Notice that we have to take care with the domain of g': it is that part of the domain of g for which g is differentiable and for which $f' \circ g(x)$ does not vanish.*

Exercise 1

Differentiate $\qquad x \longmapsto x^{1/m} \qquad (x \in R^+)$

where m is any positive integer. Compare your result with our previous result, namely

$$(x \longmapsto x^n)' = (x \longmapsto nx^{n-1}) \quad (x \in R, \quad x \neq 0)$$

where n is an integer.

8.9 Some Standard Derived Functions

The rules of differentiation obtained in the previous sections are sufficient for differentiating a rational function, that is, any function formed from polynomials by arithmetic operations (addition, multiplication, division) and by composition, and also for differentiating the inverse (if it exists) of such a function. This class of functions is quite extensive, but it does not include some very useful functions such as the trigonometric, exponential and logarithm functions. In the present section we shall fill some of the gaps. In each case it is the *result*, not the method of obtaining it, that you need to know, although we shall set some of the results as exercises so that you can get practice at applying the rules previously demonstrated. These standard derived functions are often called standard forms.

Trigonometric Functions

Denoting the derived function of the sine function by sin′, we have:

$$\sin' x = \lim_{h \to 0} \frac{\sin(x + h) - \sin x}{h}.$$

* We also assume, here and everywhere else where we differentiate a composite function, that $f \circ g$ is meaningful; i.e. that the domain of f contains the image of the domain of g.

The evaluation of this limit involves some minor technicalities, which are not an essential part of the technique of differentiation. What concerns us here is the actual value of the limit: it turns out to be simply $\cos x$, so that the formula for differentiating sine is

$$\sin' x = \cos x \quad (x \in R)$$

or simply $\sin' = \cos$.

We can also differentiate the function cos (i.e. $x \longmapsto \cos x \; (x \in R)$) using the limit definition of the derivative, but there is a possibly more elegant method which uses the identity:

$$\cos x = \sin\left(\frac{\pi}{2} - x\right) \quad (x \in R).$$

Exercise 1

Use the identity

$$\cos x = \sin\left(\frac{\pi}{2} - x\right) \quad (x \in R),$$

the composite function rule, and the fact that $\sin' = \cos$, to find the derivative of cos.

Exercise 2

Differentiate the tan function. To do this you may need to use:

(i) $\tan x = \dfrac{\sin x}{\cos x}$,

(ii) the quotient rule,

(iii) the trigonometric identity $\cos^2 x + \sin^2 x = 1$.

What is the domain of the tan function? What is the domain of its derived function?

For completeness we add the standard forms

$$\begin{aligned} \tan' &= \sec^2 \\ \sec' &= \sec \tan \\ \cot' &= -\mathrm{cosec}^2 \\ \mathrm{cosec}' &= -\mathrm{cosec}\cot \end{aligned}$$

Exponential and Logarithmic Functions

Denoting the derived function of the exponential function by exp′, we have from the definition of a derivative:

$$\exp'(x) = \lim_{h \to 0} \frac{\exp(x+h) - \exp(x)}{h} \quad (x \in R)$$

$$= \lim_{h \to 0} \frac{e^{x+h} - e^x}{h} \quad \text{by the exponential theorem}$$

$$= \lim_{h \to 0} \frac{e^x(e^h - 1)}{h}$$

$$= Ce^x$$

$$= C \exp x$$

where C denotes

$$\lim_{h \to 0} \frac{e^h - 1}{h}.$$

Now we can show that this limit exists and is equal to 1. Hence

$$\exp' = \exp. \qquad \text{Equation (1)}$$

Thus the derived function of exp is the exponential function itself.

This is a remarkable result, but it is no more than might be expected from our introduction to the exponential function via the population growth example which we considered in Chapter 5. In the model used there, the change in the world's population in a given interval of time was assumed to be proportional to the population itself.

Since differentiation gives the instantaneous rate of change, it is perhaps not so surprising that in differentiating exp we find that the derivative equals exp; we are, after all, considering a quantity (the world's population) whose instantaneous rate of change was (apart from a constant of proportionality) equal to the quantity itself.

The exponential function crops up frequently in applied mathematics because applied mathematics problems (such as the population growth problem) frequently require a function whose derived functions are related to the function itself in a way that can be reduced to Equation (1) by suitable manipulations.

Exercise 3

Use the rule for differentiating inverse functions to show that

$$\ln' : x \longmapsto \frac{1}{x} \quad (x \in R^+).$$

Exercise 4

Show that, for any real function f, with codomain R^+,

$$(\ln \circ f)' = \frac{f'}{f}.$$

Exercise 5

A useful technique, called logarithmic differentiation, for differentiating a complicated product, f, is to form the composite function, $\ln \circ f$, find its derivative and then obtain f' by using the result in Exercise 4.

(i) Apply this method to the function

$$x \longmapsto e^x \frac{(1 + x)}{1 + 2x} \quad (x \in R, \quad x \neq -\tfrac{1}{2}).$$

(ii) Use this method to differentiate $x \longmapsto x^\alpha$ $(x \in R^+)$, where α is any real number.

8.10 Additional Exercises

Exercise 1

Differentiate:

(i) $k_1 : x \longmapsto (x + 1)^5 \quad (x \in R)$

(ii) $k_2 : x \longmapsto (3x + 1)^5 \quad (x \in R)$.

Exercise 2

Differentiate:

(i) $k_1 : x \longmapsto (7x + 3)^5 \quad (x \in R)$

(ii) $k_2 : x \longmapsto (2x^2 + 3x + 2)^2 \quad (x \in R)$.

Exercise 3

Differentiate:

(i) $\sec = \dfrac{1}{\cos}$,

(ii) $\cot = \dfrac{1}{\tan}$,

(iii) $\operatorname{cosec} = \dfrac{1}{\sin}$.

The domain of the derived function is, in each case, the same as the domain of the original function. In cases (i) and (iii) the domain is R except for those numbers x for which $\cos x$ and $\sin x$ (respectively) are zero. In the case of (ii) the domain is R except for those numbers x for which $\tan x$ is undefined and for which $\tan x$ is zero. It is usual, however, to define $\cot x$ to be zero for those x for which $\tan x$ is undefined.

Exercise 4

Find $D^2(\sin)$, $D^3(\sin)$, $D^4(\sin)$, $D^5(\sin)$. Generalize your results by writing down a formula for $D^n(\sin)$ that holds for every positive integer n.

Exercise 5

Differentiate:

(i) $f:x \longmapsto x + \dfrac{1}{x} \qquad (x \in R^+)$;

(ii) $f:x \longmapsto x \exp(-x) \qquad (x \in R)$;

(iii) $f:x \longmapsto \dfrac{(x-3)^3}{(x-2)^4} \qquad (x \in R, \quad x > 3)$.

Exercise 6

(i) If

$$S:r \longmapsto 2\pi\left(r^2 + \frac{1000}{r\pi}\right) \qquad (r \in R^+)$$

find $S'(r)$.

(ii) If

$$f:x \longmapsto \sqrt{x^2 + y^2} \qquad (x \in R)$$

where y is some real number (not zero), find f'.

(iii) If

$$P(x) = (x-1)^2 \times (x+2)^2$$

defines the function P with domain R, find $P'(x)$.

8.11 Answers to Exercises

Section 8.1

Exercise 1

The position of the car at time t is given by the function

$$f: t \longmapsto at^3 \quad (t \in R_0^+, \text{ the set of all non-negative real numbers}).$$

For any t in R_0^+, the average velocity of the car over a time-interval $[t, t+h]$, $h > 0$, is given by:

$$\begin{aligned} w(t, t+h) &= \frac{a(t+h)^3 - at^3}{h} \\ &= \frac{a(t^3 + 3t^2h + 3th^2 + h^3) - at^3}{h} \\ &= 3at^2 + (3ath + ah^2) \quad (h \in R, h > 0) \end{aligned}$$

The velocity at time t is then given by

$$v(t) = \lim_{h \to 0} w(t, t+h).$$

It should be intuitively clear that $v(t) = 3at^2$ $(t \in R_0^+)$, since the terms $3ath$ and ah^2 tend to zero as h tends to zero.

Section 8.2

Exercise 1

(i) FALSE. Consider, for instance, our example of a car whose position is described by

$$f: t \longmapsto at^3 \quad (t \in R_0^+).$$

We found that

$$f'(t) = 3at^2 \quad (t \in R_0^+).$$

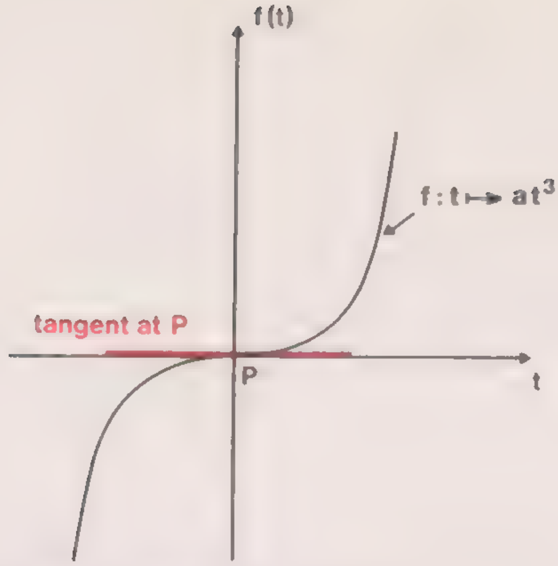
Similarly, for the function

$$g: t \longmapsto at^3 \, (t \in R)$$

we can show that

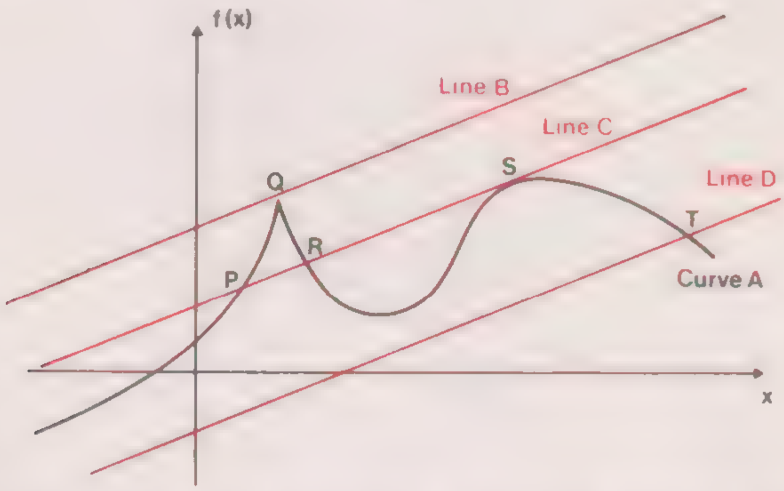
$$g'(t) = 3at^2 \quad (t \in R).$$

The derivative of g at 0 is therefore zero, so the tangent at this point is the t-axis, which crosses the curve specified by $y = g(x)$ at $(0, 0)$.



(ii) TRUE, whenever the tangent at P exists.

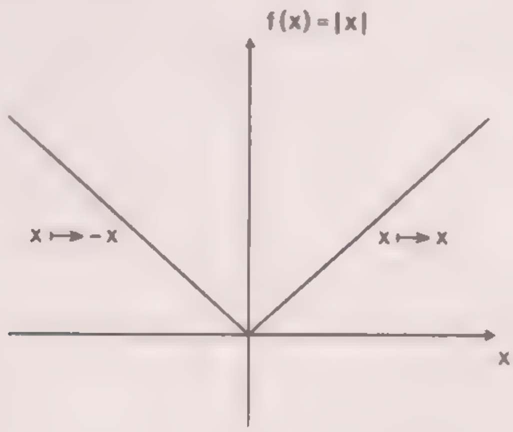
(iii) FALSE. Consider the following diagram:



Line B meets the curve A only at Q, but the tangent to the curve at Q does not exist. Line C meets the curve A at P, R and S and is a tangent to the curve at S. Line D meets the curve A at T only, but it is not the tangent to the curve at T.

(iv) FALSE. Consider the modulus function again:

$$f: x \longmapsto |x| \quad (x \in R).$$



This function is continuous at every element in its domain, and in particular at the origin, O. But

$$\frac{\Delta_h f(0)}{h} = \frac{f(0+h) - f(0)}{h} = \begin{cases} 1, \text{ if } h > 0 \\ -1, \text{ if } h < 0 \end{cases}$$

so there is no number L for which

$$\lim_{h \to 0} \frac{\Delta_h f(0)}{h} = L.$$

Exercise 2

Let $f: x \longmapsto a \ (x \in R)$ where a is any real number. Then, for all $x \in R$ and all non-zero $h \in R$, we have:

$$\begin{aligned} \frac{\Delta_h f(x)}{h} &= \frac{f(x+h) - f(x)}{h} \\ &= \frac{a - a}{h} \\ &= 0. \end{aligned}$$

Thus,

$$\begin{aligned} f'(x) &= \lim_{h \to 0} 0 \\ &= 0. \end{aligned}$$

Exercise 3

For all $t \in R$ and all non-zero $h \in R$, we have:

$$\begin{aligned} \frac{\Delta_h f(t)}{h} &= \frac{f(t+h) - f(t)}{h} \\ &= \frac{(t+h)^2 - t^2}{h} \\ &= 2t + h. \end{aligned}$$

Thus,

$$\begin{aligned} f'(t) &= \lim_{h \to 0} (2t + h) \\ &= 2t. \end{aligned}$$

In particular,

$$f'(-3) = -6,$$
$$f'(0) = 0,$$
$$f'(2) = 4;$$

so the slope of the tangent of the graph is negative at -3, zero at 0 and positive at 2.

Section 8.3

Exercise 1

We must differentiate the function

$$f: x \longmapsto ax^2 + bx + c \quad (x \in R).$$

That is, we must find

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{a(x+h)^2 + b(x+h) + c - (ax^2 + bx + c)}{h} \\ &= \lim_{h \to 0} \frac{2axh + ah^2 + bh}{h} \\ &= \lim_{h \to 0} (2ax + ah + b) \\ &= 2ax + b. \end{aligned}$$

The derived function is, therefore:

$$f': x \longmapsto 2ax + b \quad (x \in R).$$

Notice that if we write

$$f_1 : x \longmapsto ax^2, f_2 : x \longmapsto bx, f_3 : x \longmapsto c \quad (x \in R),$$

then

$$f = f_1 + f_2 + f_3$$

and

$$f' = f'_1 + f'_2 + f'_3.$$

Exercise 2

$$f'(x) = \lim_{h \to 0} \frac{(x + h)^n - x^n}{h}$$

$$= \lim_{h \to 0} \frac{\left\{x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \cdots + h^n\right\} - x^n}{h}$$

(by the binomial theorem)

$$= \lim_{h \to 0} \left\{nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}h + \cdots + h^{n-1}\right\}$$

(on dividing by h, since $h \neq 0$)

$$= nx^{n-1}$$

Consequently the derived function of

$$x \longmapsto x^n \quad (x \in R)$$

is

$$x \longmapsto nx^{n-1} \quad (x \in R)$$

Exercise 3

(i) When $t \in$ (domain of f') then

$$\lim_{h \to 0} \left(\frac{af(t + h) - af(t)}{h}\right) = \lim_{h \to 0} \left(a\left[\frac{f(t + h) - f(t)}{h}\right]\right)$$

$$= \lim_{h \to 0} (a) \times \lim_{h \to 0} \left(\frac{f(t + h) - f(t)}{h}\right)$$

$$= af'(t).$$

(ii) When $t \in$ (intersection of domains of f' and g'), then

$$\lim_{h \to 0} \left(\frac{\{f(t + h) + g(t + h)\} - \{(f(t) + g(t)\}}{h}\right)$$

$$= \lim_{h \to 0} \left(\frac{f(t + h) - f(t)}{h} + \frac{g(t + h) - g(t)}{h}\right)$$

$$= \lim_{h \to 0} \left(\frac{f(t + h) - f(t)}{h}\right) + \lim_{h \to 0} \left(\frac{g(t + h) - g(t)}{h}\right)$$

$$= f'(t) + g'(t).$$

Exercise 4

(i) Using the formula for the derivative of a polynomial function, the derived function of

$$x \longmapsto 10x^5 + \tfrac{1}{8}x^3 + x \quad (x \in R)$$

is

$$x \longmapsto 50x^4 + \tfrac{3}{8}x^2 + 1 \quad (x \in R).$$

(ii) Let $f: x \longmapsto 2x$, and $g: x \longmapsto |x|$ $(x \in R)$. Then $f': x \longmapsto 2$ $(x \in R)$ and

$$\left.\begin{array}{ll} g': x \longmapsto 1 & (x > 0) \\ x \longmapsto -1 & (x < 0) \end{array}\right\} \quad (x \in R, \quad x \neq 0)$$

So the derived function of $x \longmapsto 2x + |x|$ $(x \in R)$ is

$$x \longmapsto f'(x) + g'(x) \quad (x \in R, \quad x \neq 0)$$

i.e.

$$x \longmapsto \left\{\begin{array}{ll} 3 & (x > 0) \\ 1 & (x < 0) \end{array}\right\} \quad (x \in R, \quad x \neq 0).$$

Section 8.4

Exercise 1

(i) Let

$$f: x \longmapsto ax^2 + bx + c \quad (x \in R)$$

then

$$Df: x \longmapsto 2ax + b \quad (x \in R)$$

and

$$D^2f: x \longmapsto 2a \quad (x \in R).$$

(ii)

$$f: x \longmapsto x^3 \quad (x \in R)$$

$$Df: x \longmapsto 3x^2 \quad (x \in R)$$

$$D^2f: x \longmapsto 6x \quad (x \in R)$$

$$D^3f: x \longmapsto 6 \quad (x \in R).$$

(Note the fact that we have included the somewhat trivial information $(x \in R)$ on every line, to emphasize that the domain stays the same throughout. It looks rather unnecessary here, but if we were differentiating more

complicated functions we might well find that the domain changed in going from a function to its derived function.)

Section 8.5

Exercise 1

A suitable simple example is:

$$f : x \longmapsto x \quad (x \in R), \qquad Df : x \longmapsto 1 \quad (x \in R)$$

and

$$g : x \longmapsto 1 \quad (x \in R), \qquad Dg : x \longmapsto 0 \quad (x \in R);$$

then

$$f \times g : x \longmapsto x \quad (x \in R), \qquad D(f \times g) : x \longmapsto 1 \quad (x \in R).$$

But

$$\begin{aligned} Df \times Dg &= x \longmapsto 0 \quad (x \in R) \\ &\neq D(f \times g). \end{aligned}$$

Exercise 2

$$fg : x \longmapsto 2x^2 - x - 1 \qquad (x \in R)$$

$$(fg)' : x \longmapsto 4x - 1 \qquad (x \in R)$$

$$\begin{aligned} f(x)g'(x) + g(x)f'(x) &= (2x + 1)g'(x) + (x - 1)f'(x) \\ &= (2x + 1)1 + (x - 1)2 \\ &= 4x - 1. \end{aligned}$$

Thus

$$(fg' + gf') : x \longmapsto 4x - 1 \qquad (x \in R),$$

and the result is verified.

Section 8.6

Exercise 1

Almost any example will do; here is one:

Let

$$g : x \longmapsto 1 \quad (x \in R), \qquad g' : x \longmapsto 0 \quad (x \in R);$$

$$f : x \longmapsto x \quad (x \in R), \qquad f' : x \longmapsto 1 \quad (x \in R);$$

$$f \circ g : x \longmapsto 1 \quad (x \in R);$$

$$(f \circ g)' : x \longmapsto 0 \quad (x \in R), \qquad f' \circ g' : x \longmapsto 1 \quad (x \in R).$$

So $D(f \circ g) \neq Df \circ Dg$ in this case.

Exercise 2

We have:

$$f \circ g : x \longmapsto (x^2 - 1)^2 + 1 = x^4 - 2x^2 + 2 \quad (x \in R)$$

giving by direct differentiation:

$$(f \circ g)' : x \longmapsto 4x^3 - 4x \quad (x \in R).$$

Using the composite function rule we have:

$$(f \circ g)' = (f' \circ g) \times g'$$

where

$$f' : x \longmapsto 2x \quad (x \in R)$$

and

$$g' : x \longmapsto 2x \quad (x \in R)$$

Thus

$$(f \circ g)' : x \longmapsto [2(x^2 - 1)] \times 2x = 4x^3 - 4x \quad (x \in R).$$

As we would expect, the results are the same.

Section 8.7

Exercise 1

$$\left(\frac{u}{v}\right)' = [u \times (r \circ v)]' = u' \times (r \circ v) + u \times (r \circ v)' \qquad \text{by (ii)}$$

$$= u' \times (r \circ v) + u \times [(r' \circ v) \times v'] \qquad \text{by (iii)}$$

$$= u' \times (r \circ v) + u \times (-r^2 \circ v) \times v'.$$

Now $r \circ v = \dfrac{1}{v}$ and $-r^2 \circ v = \dfrac{-1}{v^2}$, so that

$$\left(\frac{u}{v}\right)' = \frac{u'}{v} - \frac{u \times v'}{v^2}$$

$$= \frac{u' \times v - v' \times u}{v^2}.$$

Exercise 2

There are many ways of tackling this: we shall use the quotient rule with

$$u : x \longmapsto 1 \qquad (x \in R, \quad x \neq 0)$$

$$v : x \longmapsto x^k \qquad (x \in R, \quad x \neq 0)$$

We already know that

$$v' : x \longmapsto kx^{k-1} \quad (x \in R, \quad x \neq 0).$$

so that

$$\frac{u'(x) \times v(x) - u(x) \times v'(x)}{v^2} = \frac{0 \times x^k - 1 \times kx^{k-1}}{x^{2k}}$$

$$= -kx^{-k-1}$$

$$= \frac{-k}{x^{k+1}}.$$

Considering the form $-kx^{-k-1}$, and remembering that we were differentiating $x \longmapsto x^{-k}$ $(x \in R,\ x \neq 0)$, we can see that we can now write

$$D(x \longmapsto x^n) = x \longmapsto nx^{n-1}$$

for any integral value of n,
(the domain is the set of non-zero real numbers when n is negative).

Exercise 3

The answer to both parts is

$$x \longmapsto \frac{4}{(2x+1)^2} \qquad (x \in R, \quad x \neq -\tfrac{1}{2}).$$

The two results are the same because $w_1 - w_2$ is a constant function:

$$w_2(x) - w_1(x) = \frac{2x-1}{2x+1} + \frac{2}{2x+1}$$

$$= \frac{(2x+1)-2}{2x+1} + \frac{2}{2x+1}$$

$$= 1$$

so that

$$w_2 - w_2 = x \longmapsto 1 \qquad (x \in R, \quad x \neq -\tfrac{1}{2}).$$

whence

$$(w_2 - w_1)' = w_2' - w_1' = (x \longmapsto 1)' = x \longmapsto 0$$

i.e.

$$w_2'(x) - w_1'(x) = 0$$

or

$$w_2'(x) = w_1'(x).$$

Section 8.8

Exercise 1

Call the given function g. Then the inverse, f, of g is given by

$$f : x \longmapsto x^m \qquad (x \in R^+),$$

so that

$$f' : x \longmapsto mx^{m-1} \quad (x \in R^+).$$

The inverse function rule then gives

$$\begin{aligned} g'(x) &= \frac{1}{m(x^{1/m})^{m-1}} \\ &= \frac{1}{m} \times \frac{1}{x^{(1-1/m)}} \\ &= \frac{1}{m} x^{1/m-1} \end{aligned}$$

i.e.

$$g' : x \longmapsto \frac{1}{m} x^{1/m-1}$$

We have shown that the result:

the derivative of $x \longmapsto x^n$ is $x \longmapsto nx^{n-1}$, where $n \in Z$, also holds when n is the reciprocal of a positive integer.

Section 8.9

Exercise 1

Let

$$g : x \longmapsto \frac{\pi}{2} - x \quad (x \in R)$$

$$f : x \longmapsto \sin x \qquad (x \in R).$$

Then using $(f \circ g)' = (f' \circ g) \times g'$, we have

$$\cos' = \left[\cos \circ \left(x \longmapsto \frac{\pi}{2} - x\right)\right] \times (x \longmapsto -1)$$

i.e.

$$\cos' x = -\cos\left(\frac{\pi}{2} - x\right).$$

This result can be simplified by noting that $\cos\left(\frac{\pi}{2} - x\right) = \sin x \ (x \in R)$, so that

$$\cos' = -\sin.$$

Exercise 2

Starting with

$$\tan x = \frac{\sin x}{\cos x}$$

we have

$$\begin{aligned}\tan' x &= \frac{\sin' x \cos x - \cos' x \sin x}{\cos^2 x} && \text{(quotient rule)}\\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x}\\ &= \frac{1}{\cos^2 x} && (\cos^2 x + \sin^2 x = 1)\\ &= \sec^2 x && \left(\sec x = \frac{1}{\cos x}\right).\end{aligned}$$

This result can be written as

$$\tan' = \sec^2.$$

The domain of tan is R except for

$$\ldots, \frac{-5}{2}\pi, \frac{-3}{2}\pi, \frac{-1}{2}\pi, \frac{1}{2}\pi, \frac{3}{2}\pi, \frac{5}{2}\pi, \ldots.$$

The derived function has the same domain.

Exercise 3

Since ln is the inverse function of exp, the inverse function rule gives

$$\ln'(x) = \frac{1}{\exp'(\ln x)} = \frac{1}{\exp(\ln x)} = \frac{1}{x} \quad (x \in R^+).$$

Exercise 4

The composite function rule gives

$$(\ln \circ f)' = (\ln' \circ f) \times f' = \frac{f'}{f}.$$

Exercise 5

(i) Let

$$f(x) = e^x\left(\frac{1+x}{1+2x}\right);$$

then

$$\ln f(x) = x + \ln(1+x) - \ln(1+2x)$$

$$\therefore\ \frac{f'(x)}{f(x)} = 1 + \frac{1}{1+x} - \frac{2}{1+2x} \quad \text{on differentiating}$$

$$\therefore\ f'(x) = f(x)\left(1 + \frac{1}{1+x} - \frac{2}{1+2x}\right)$$

$$= e^x\left(\frac{1+x}{1+2x}\right)\left(1 + \frac{1}{1+x} - \frac{2}{1+2x}\right).$$

(ii) Let

$$f: x \longmapsto x^\alpha \quad (x \in R^+)$$

$$f(x) = x^\alpha$$

$$\ln f(x) = \ln(x^\alpha) = \alpha \ln x.$$

Differentiating gives

$$\frac{f'(x)}{f(x)} = \alpha\frac{1}{x}$$

$$f'(x) = \alpha\frac{1}{x}f(x) = \frac{\alpha}{x}x^\alpha = \alpha x^{\alpha-1}$$

i.e.

$$f': x \longmapsto \alpha x^{\alpha-1} \quad (x \in R^+).$$

We have therefore shown that the function

$$f: x \longmapsto x^\alpha \qquad (x \in R^+)$$

has derived function

$$f': x \longmapsto \alpha x^{\alpha-1} \qquad (x \in R^+)$$

where α is any real number.

Section 8.10

Exercise 1

(i) Let

$$f_1 : x \longmapsto x^5 \qquad (x \in R)$$

$$g_1 : x \longmapsto x + 1 \quad (x \in R).$$

Thus

$$f_1 \circ g_1 : x \longmapsto (x + 1)^5 \quad (x \in R)$$

Using

$$(f \circ g)' = (f' \circ g) \times g'$$

we obtain

$$k_1' : x \longmapsto 5(x + 1)^4$$

(ii) We can repeat the process with

$$f_2 : x \longmapsto x^5 \qquad (x \in R)$$

$$g_2 : x \longmapsto 3x + 1 \quad (x \in R).$$

Alternatively, we notice that

$$k_2(x) = k_1(3x)$$

so that we can write

$$k_2 = k_1 \circ (x \longmapsto 3x)$$

whence

$$\begin{aligned} k_2' &= [k_1' \circ (x \longmapsto 3x)] \times (x \longmapsto 3) \\ &= [x \longmapsto 5(3x + 1)^4] \times (x \longmapsto 3) \\ &= x \longmapsto 15(3x + 1)^4 \quad (x \in R). \end{aligned}$$

In general, if

$$k_2(x) = k_1(ax)$$

so that

$$k_2 = k_1 \circ (x \longmapsto ax)$$

we have

$$k_2' = [k_1' \circ (x \longmapsto ax)] \times (x \longmapsto a)$$

so that

$$k_2'(x) = ak_1'(ax).$$

Exercise 2

(i) $x \longmapsto 35(7x+3)^4 \quad (x \in R)$

(ii) $x \longmapsto (8x+6)(2x^2+3x+2) \quad (x \in R).$

Exercise 3

The rule for differentiating quotients of the form:

$$u(x) = \frac{v(x)}{w(x)} \quad (x \in R, \text{ where } w(x) \neq 0)$$

is

$$u'(x) = \frac{w(x)v'(x) - v(x)w'(x)}{(w(x))^2}$$

If $v(x) = 1$, $v'(x) = 0$, so that the quotient rule becomes:

$$u'(x) = \frac{-w'(x)}{(w(x))^2}$$

(i) When $u(x) = \dfrac{1}{\cos x}$, then $w(x) = \cos x$ and $w'(x) = -\sin x$, so that

$$u'(x) = \frac{\sin x}{\cos^2 x} = \sec x \tan x$$

i.e.

$$\sec' = \sec \tan$$

(ii) When $u(x) = \dfrac{1}{\tan x}$, then $w(x) = \tan x$ and $w'(x) = \sec^2 x$, so that

$$u'(x) = -\frac{\sec^2 x}{\tan^2 x} = -\frac{1}{\cos^2 x} \times \frac{\cos^2 x}{\sin^2 x} = \frac{-1}{\sin^2 x} = -\operatorname{cosec}^2 x$$

i.e.

$$\cot' = -\operatorname{cosec}^2$$

(iii) $u(x) = \dfrac{1}{\sin x}$; $w(x) = \sin x$; and so $w'(x) = \cos x$. Thus

$$u'(x) = -\frac{\cos x}{\sin^2 x} = -\operatorname{cosec} x \cot x$$

i.e.

$$\operatorname{cosec}' = -\operatorname{cosec} \cot.$$

Exercise 4

$$D \sin = \cos$$

$$D^2 \sin = D \cos = -\sin$$

$$D^3 \sin = D(D^2 \sin) = D(-\sin) = -\cos$$

$$D^4 \sin = D(D^3 \sin) = D(-\cos) = \sin$$

$$D^5 \sin = D(D^4 \sin) = D(\sin) = \cos.$$

Suppose n is an *odd* integer i.e. we can write

$$n = 2k + 1 \quad \text{where } k \in \{0, 1, 2, \ldots\},$$

then $\quad D^{2k+1} \sin = (-1)^k \cos.$

If n is an *even* integer i.e. we can write

$$n = 2k, \quad (k \in \{1, 2, 3, \ldots\})$$

then $\quad D^{2k} \sin = (-1)^k \sin.$

Exercise 5

(i) Standard derived functions give

$$(x \longmapsto x)' = (x \longmapsto 1),$$

and

$$(x \longmapsto x^{-1})' = (x \longmapsto -x^{-2});$$

it follows by the addition rule that

$$\left(x \longmapsto x + \frac{1}{x}\right)' = \left(x \longmapsto 1 - \frac{1}{x^2}\right) \qquad (x \in R^+).$$

(ii) The given function has the product form

$$f = gh$$

where

$$g(x) = x \qquad (x \in R)$$

$$h(x) = \exp(-x) \quad (x \in R).$$

Moreover, $h(x)$ is of the form $\exp(k(x))$ where

$$k(x) = -x \qquad (x \in R).$$

By the "function of a function" rule, the derived function of h is given by

$$\begin{aligned} h'(x) &= \exp'(k(x)) \times k'(x) = \exp(-x)(-1) \\ &= -\exp(-x). \end{aligned}$$

By the product rule, the derived function of f is therefore given by:

$$\begin{aligned} f'(x) &= g'(x)h(x) + g(x)h'(x) \\ &= 1\exp(-x) + x(-\exp(-x)) \\ &= (1 - x)\exp(-x) \qquad (x \in R). \end{aligned}$$

(iii) We could differentiate directly using the quotient rule, but it is easier to use logarithmic differentiation (see Exercise 8.9.5). That is, we write down the formula for $f(x)$ as usual

$$f(x) = \frac{(x-3)^3}{(x-2)^4}$$

but before differentiating we take natural logarithms, obtaining

$$\ln f(x) = 3\ln(x-3) - 4\ln(x-2)$$

and then, by the "function of a function" rule,

$$\ln'(f(x))f'(x) = 3\ln'(x-3) - 4\ln'(x-2)$$

(since $(x \longmapsto x - 3)' = x \longmapsto 1$ and $(x \longmapsto x - 2)' = x \longmapsto 1$).

Since $\ln'(u) = \frac{1}{u}$, this simplifies to

$$\frac{f'(x)}{f(x)} = \frac{3}{x-3} - \frac{4}{x-2}$$

so that

$$\begin{aligned} f'(x) &= \left(\frac{3}{x-3} - \frac{4}{x-2}\right) f(x) \\ &= \left(\frac{3}{x-3} - \frac{4}{x-2}\right)\frac{(x-3)^3}{(x-2)^4} \qquad (x \in R,\ x > 3). \end{aligned}$$

Exercise 6

(i) $$S(r) = 2\pi\left(r^2 + \frac{1000}{\pi r}\right).$$

Standard derived functions and the rule for constant factors give

$$(r \longmapsto r^2)' = (r \longmapsto 2r),$$

and

$$\left(r \longmapsto \frac{1000}{\pi} r^{-1}\right)' = \left(r \longmapsto -\frac{1000}{\pi} r^{-2}\right),$$

therefore the addition rule and the rule for multiplication by a constant give

$$S'(r) = 2\pi\left(2r - \frac{1000}{\pi r^2}\right).$$

(ii) The function can be expressed as a composition:

$$f(x) = g(h(x))$$

where

$$h(x) = x^2 + y^2 \quad (x \in R)$$

and

$$g(u) = \sqrt{u} \qquad (u \in R^+).$$

The derived functions of h and g are such that

$$h'(x) = 2x + 0 \quad \text{(by the sum rule)}$$

$$g'(u) = \tfrac{1}{2}u^{-1/2} = \frac{1}{2\sqrt{u}}.$$

The rule for differentiating composite functions gives

$$f'(x) = g'(h(x)) \times h'(x)$$

$$= \frac{1}{2\sqrt{h(x)}} \times 2x$$

$$= \frac{x}{\sqrt{x^2 + y^2}}.$$

and so the answer to the exercise is

$$f' : x \longmapsto -\frac{x}{\sqrt{x^2 + y^2}} \quad (x \in R).$$

(Since y is not zero, the denominator of $f'(x)$ is never zero.)

(iii) The given function is of the form

$$P = QS$$

with

$$Q(x) = (x - 1)^2 \quad (x \in R)$$

$$S(x) = (x + 2)^2 \quad (x \in R).$$

These functions are compositions; for example

$$Q = F \circ G$$

where

$$F(u) = u^2 \qquad (u \in R)$$

$$G(x) = x - 1 \qquad (x \in R).$$

The rule for differentiating composite functions gives

$$\begin{aligned} Q'(x) &= F'(G(x)) \times G'(x) \\ &= 2G(x) \times 1 \\ &= 2(x - 1). \end{aligned}$$

Similarly, we have $S'(x) = 2(x + 2)$.

The product rule therefore gives

$$\begin{aligned} P'(x) &= Q'(x)S(x) + Q(x)S'(x) \\ &= 2(x - 1)(x + 2)^2 + (x - 1)^2 2(x + 2) \\ &= 2(x - 1)(x + 2)(2x + 1). \end{aligned}$$

Appendix : The Leibniz Notation

Probably the most widely used notation for calculus is the one invented by Leibniz. In this notation, if x is a variable representing an element in the domain of a function, f, and y is another variable whose value is

related to that of x by

$$y = f(x) \qquad (x \in \text{domain of } f),$$

then we use the symbol $\frac{dy}{dx}$ to stand for $f'(x)$. That is, we define

$$\frac{dy}{dx} = f'(x) \quad (x \in \text{domain of } f').$$

The advantage of the Leibniz notation is its conciseness; for example, the formula giving the derived function of $x \longmapsto x^2$ ($x \in R$) can be written

$$\frac{d(x^2)}{dx} = 2x$$

instead of

$$(x \longmapsto x^2)' = (x \longmapsto 2x).$$

The process of calculating $\frac{dy}{dx}$, when y is defined as an expression involving x (for example the x^2 in the above example), is called *differentiating* this expression *with respect to* x. Formally, we may say that to differentiate y with respect to x is to calculate $f'(x)$ where f is the function defined by

$$f: x \longmapsto y.$$

The disadvantage of the Leibniz notation is that it contains some traps for the beginner, arising from the difficulty of assigning an independent meaning to the symbols dx and dy when they occur separately (rather than locked together in the combination $\frac{dy}{dx}$).

Thus the rules for differentiating sums and products in the Leibniz notation become

$$\frac{d(u+v)}{dx} = \frac{du}{dx} + \frac{dv}{dx},$$

$$\frac{d(cu)}{dx} = c\frac{du}{dx} \text{ if } c \text{ is a constant}$$

(i.e. if $x \longmapsto c$ is a constant function),

$$\frac{d(uv)}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}.$$

To express the rule for differentiating composite functions in this notation,

let x, y and z be variables related by

$$y = g(x) \qquad (x \in \text{domain of } g)$$

$$z = f(y) \qquad (y \in \text{domain of } f)$$

so that

$$g'(x) = \frac{dy}{dx} \qquad (x \in \text{domain of } g'),$$

$$f'(g(x)) = f'(y) = \frac{dz}{dy} \qquad (y \in \text{domain of } f'),$$

Then the rule for differentiating a function h defined by

$$h(x) = f(g(x)) \qquad (x \in \text{domain of } h)$$

is

$$h'(x) = f'(g(x)) \times g'(x) \quad (x \in \text{domain of } h')$$

which takes the simple form

$$\frac{dz}{dx} = \frac{dz}{dy} \times \frac{dy}{dx}.$$

This rule is called the chain rule and is easy to remember because one can think of dy as "cancelling out" from the expression on the right. The rule for differentiating inverse functions also takes a convenient form in this notation: it is

$$\frac{dy}{dx} = \frac{1}{\left(\frac{dx}{dy}\right)}.$$

The proof is similar to the proof of the chain rule. We shall not go into details here.

CHAPTER 9 THE FUNDAMENTAL THEOREM OF CALCULUS

9.0 Introduction

In Chapters 7 and 8 we have discussed *integration* and *differentiation*. You may have already suspected, or have been told, that there is a connection between them. In this chapter we shall investigate this connection.

We begin by looking back to the *definite integral* which we discussed in Chapter 7, and we introduce the concept of *primitive function*.

We then introduce the *fundamental theorem of calculus* in two parts, first showing that differentiation undoes integration, and then that integration undoes differentiation.

This concludes the first volume on calculus, leaving further techniques of calculus and its application to be developed in the next volume.

9.1 Primitive Functions

At first sight there seems to be no connection at all between differentiation and integration, or between tangents to curves and the areas under the curves. The first step in seeing that there actually *is* a connection is to look more closely at the structure of the formulas for integrals obtained earlier in the course.

In Chapter 7 we saw that one way of looking at the definite integral is as an area, and by approximating areas by sums of rectangles we were able to find exact formulas for a few definite integrals, for example:

$$\int_a^b x \longmapsto 1 = b - a,$$

$$\int_a^b x \longmapsto x = \frac{b^2}{2} - \frac{a^2}{2},$$

$$\int_a^b x \longmapsto x^2 = \frac{b^3}{3} - \frac{a^3}{3},$$

where a and b are real numbers, and the various functions are all real functions (that is, with domain and codomain R or a subset of R; the domain must, of course, include the interval $[a, b]$). Although the expressions on the right-hand sides of the three equations are all different, they have a common feature: each of them is the difference of two terms, one depending on b and the other depending *in the same way* on a. Let us use the letter F to denote the real function which specifies the way in

which the first term on the right depends on b (for example, $F: b \longmapsto \frac{b^2}{2}$ in the second formula); then the first term on the right is $F(b)$, and the second is $F(a)$, and each of the formulas can be written in the form:

$$\int_a^b f = F(b) - F(a) \qquad \textbf{Equation (1)}$$

with suitable functions f and F. We shall call the function F a primitive function* of the function f; in our third example, $b \longmapsto \frac{b^3}{3}$ is a primitive function of $x \longmapsto x^2$. In general, given any continuous† real function f, we define a primitive function of f to be any F such that Equation (1) holds for all a and b in the domain of f.

Notice that we say "*a* primitive function", not "*the* primitive function". This is because primitive functions are not unique: for each f there are *many* primitive functions F. For instance, instead of

$$F: b \longmapsto \frac{b^2}{2}$$

in the second example above, we could choose

$$F_1: b \longmapsto \frac{b^2}{2} + 3$$

and still have

$$\int_a^b x \longmapsto x^2 = F_1(b) - F_1(a),$$

and so F_1 is also a primitive function of f.

Exercise 1

If f is a continuous real function with a primitive function F, use Equation (1) to show that

$$\int_a^b f = -\int_b^a f$$

for all a and b in the domain of f.

* The term indefinite integral is common. Our terminology is meant to emphasize that F is a *function*, not a number like a definite integral.

† For the definition of continuity, see Chapter 4, Section 4.2.

Exercise 2

Give the values of $F(b)$ and $F(a)$, and the functions F, which will complete the following table:

f	$F(b)$	$F(a)$	F, a primitive function of f
$x \longmapsto 1 \quad (x \in R)$ $x \longmapsto x \quad (x \in R)$ $x \longmapsto x^2 \quad (x \in R)$			

For a given continuous real function f with domain R, the definite integral $\int_a^b f$ is determined by the values of both a and b; evaluating it is therefore tantamount to calculating the image of (a, b) under the following function of two variables:

$$(a, b) \longmapsto \int_a^b f \qquad ((a, b) \in R \times R).$$

If we regard either a or b as fixed, then we can consider the definite integral as defining a function of one variable, with domain R rather than $R \times R$: for example,

$$b \longmapsto \int_a^b f \qquad (b \in R).$$

Functions with domain R are usually easier to deal with than those with domain $R \times R$, but it is not clear at the moment just how this new function is going to help. This is where the Fundamental Theorem of Calculus comes in: it gives us a general method for finding a primitive function of f *without* first evaluating the integral by summing rectangles.

The first and last columns in Exercise 2 constitute a list of ordered pairs of functions, and may therefore be held to define a mapping whose domain and codomain are sets of functions — that is, an *operator*. There is no need to restrict the domain of this operator to the three functions listed in Exercise 2; rather, we may expect to be able to use for the domain some much more general set of functions f, such that $\int_a^b f$ exists for all a and b in the domain of f. This operator deserves a name. It is called

the integration operator, and represented symbolically as follows:

$$I: f \longmapsto \text{(the set of all primitive functions of } f\text{)}.$$

The process of finding a primitive function is called integration, and in applying I to f one is said to integrate the function f.

Exercise 3

If f is a real continuous function with domain R, and F is a primitive function of f, are there any numbers c (other than zero) for which the function F_c defined by

$$F_c : x \longmapsto F(x) + c \qquad (x \in R)$$

is also a primitive function of f?

The result of this last exercise is an important one. If F is any primitive function of some given function f, then any function of the form

$$x \longmapsto F(x) + c \qquad (x \in \text{domain of } F)$$

where c is any real number, is also a primitive function of f. Another way of saying the same thing is that the operator I is not a function: under this operator the image $I(f)$ of a given element f in the domain of I is not a *unique* element of the codomain of I, but a *set* of such elements. The real number c is called a constant of integration, and each different value for c gives a different primitive function of f.

Exercise 4

Find a primitive function F of the function

$$x \longmapsto x \qquad (x \in R)$$

with the property $F(0) = 1$.

9.2 The Fundamental Theorem: Part 1

If the only property of primitive functions F of a function f were that they satisfied the definition given in the preceding section; that is,

$$\int_a^b f = F(b) - F(a) \qquad ((a, b) \in R \times R), \qquad \textbf{Equation (1)}$$

they would provide little more than a useful alternative notation for definite integrals. They would not help us with the job of actually calculating the definite integrals, because our only way of finding a primitive

function of f would be to find the definite integral first, and then use Equation (1) to find F, by taking the terms in either b or a. The property which makes the primitive function concept really useful is that there is another way of finding primitive functions, which does not require us to find the corresponding definite integral first. This method is provided by the Fundamental Theorem of Calculus.

The basic idea of the Fundamental Theorem is to creep up on the integration mapping from behind, as it were, by identifying its reverse mapping. In view of the property we used to define I,

$$I : f \longrightarrow F,$$

this is equivalent to finding a rule giving f in terms of one of its primitives. In other words, we shall regard F (rather than f) as the given function in Equation (1), and try to determine from it the function f. This will enable us to identify the mapping

$$F \longmapsto f;$$

we can then find primitives of f by reversing this new mapping instead of by evaluating definite integrals directly.

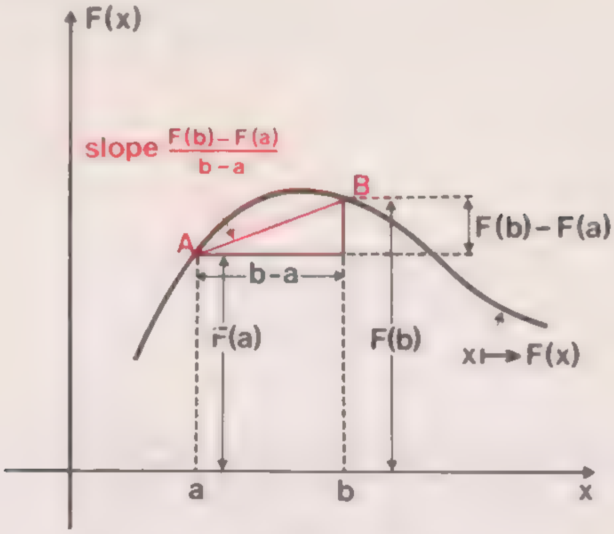
We assume as usual that f and F are real functions, and we shall also assume, in order to be able to state theorems that can be proved rigorously (even though we do not prove them rigorously here), that f is continuous everywhere in its domain. Let us consider a related problem since there is no immediately obvious way to go about finding f from F using Equation (1). We know that $\int_a^b f$ is the area under the graph of $f(x)$. If we divide this area by $(b - a)$ we shall obtain the average value of f over the interval. $[a, b]$. We can thus write

$$f_{av}[a, b] = \frac{1}{b-a}\int_a^b f.$$

Equation (1) now gives us this average value in terms of F:

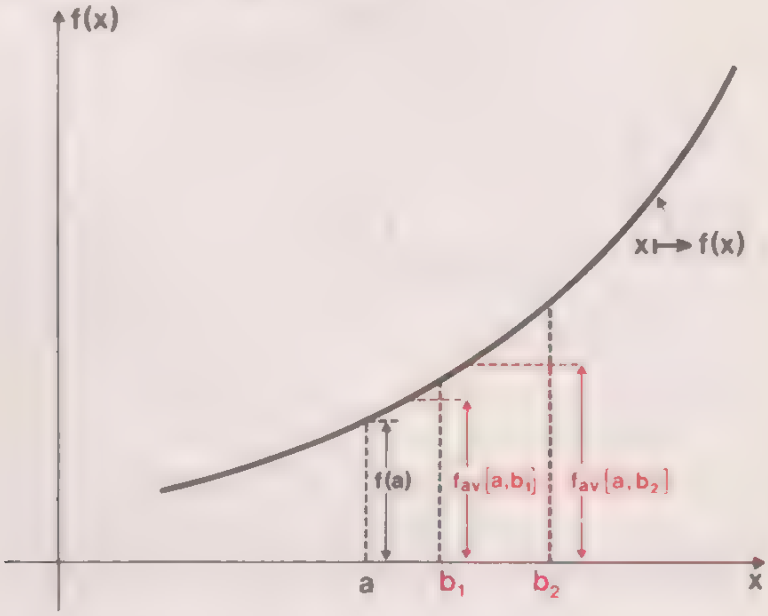
$$f_{av}[a, b] = \frac{F(b) - F(a)}{b - a}. \qquad \text{Equation (2)}$$

The expression on the right can be interpreted graphically: it is the slope of a chord of the graph of F, as illustrated in the following figure:



Finding the average value of f(x) from the graph of F

We now need a way to get from the *average* value of f over $[a, b]$ to the value of f *at some specific point* in its domain. We had a problem in Chapter 8 when we wanted to obtain *instantaneous* velocities from *average* velocities, and the method is just the same here. To refresh your memory, we shall go quickly through the argument again. Since f is continuous, we can argue that, if a and b are very close together, then $f(x)$ is very nearly constant over the interval $[a, b]$, so that its value at any specific point in the interval, say a, is closely approximated by the average value over $[a, b]$, as illustrated in the figure below.



How the average value of f(x) varies as the interval width is reduced

By making b close enough to a we expect to make the error in this approximation as small as we please, and by a suitable limiting procedure we

expect to obtain $f(a)$ exactly. To formulate this idea more precisely, we expect to find that

$$\lim_{b \to a} (f_{av}[a, b]) = f(a), \qquad \text{Equation (3)}$$

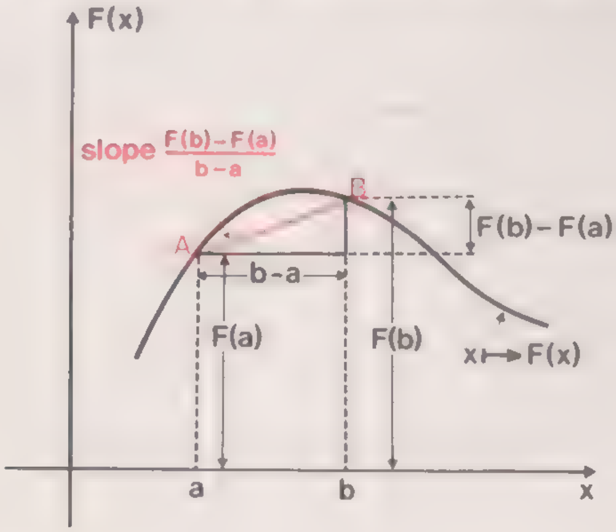
that is

$$\lim_{b \to a} \left(\frac{1}{b - a} \int_a^b f \right) = f(a).$$

It can be shown that this equation does indeed hold (and that the limit on the left exists), when the function f is continuous at a. Substituting from Equation (2) for the left-hand side of Equation (3), we find

$$\lim_{b \to a} \frac{F(b) - F(a)}{b - a} = f(a).$$

In terms of the graph of F, the limit on the left is the limiting slope of the chord AB when B is very close to A. We already know from Chapter 8 that this is the slope of the tangent at A given by the derivative of F at a, that is, by $DF(a)$, where D is the differentiation operator.



Thus we have found that

$$DF(a) = f(a).$$

Since this equation holds for any real a in the domain of f, it follows that the functions DF and f are identical. This is the first part of the Fundamental Theorem of Calculus:

> If f is a real continuous function and if F is a primitive function of f, then $DF = f$.

In other words, the differentiation which takes us from F to f *undoes* the integration which took us from f to F.

Example 1

We have already seen that a primitive function of $x \longmapsto x$ is $x \longmapsto \frac{1}{2}x^2$. How does this fit in with the theorem?

In the context of the theorem, $x \longmapsto x$ is f and $x \longmapsto \frac{1}{2}x^2$ is F. According to the theorem, $DF = f$, and indeed we see here that

$$D(x \longmapsto \tfrac{1}{2}x^2) = x \longmapsto x.$$

Exercise 1

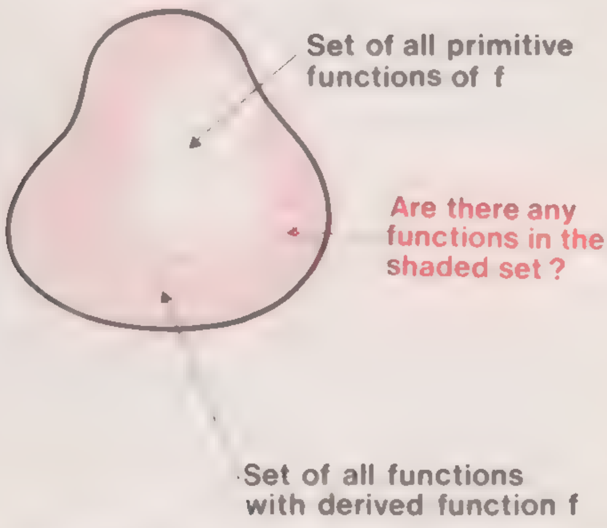
Use the Fundamental Theorem to check whether the following statements are true or false:

(i) $\int_a^b x \longmapsto \sin x = \cos b - \cos a$

(ii) $\int_a^b x \longmapsto \cos x = \sin b - \sin a.$

The first part of the Fundamental Theorem of Calculus does not completely solve the problem of finding primitive functions, but it takes us a long way towards the solution. It does not show us how to *find* a primitive function, F, of a given continuous function f. It *does* tell us that each primitive function F can be differentiated, and has derived function f. We can use the theorem to find f when F is known, or to check the calculation by which a primitive function has been found.

9.3 The Fundamental Theorem: Part 2

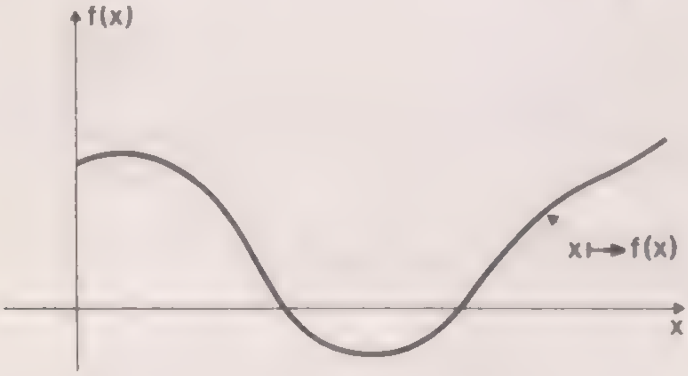


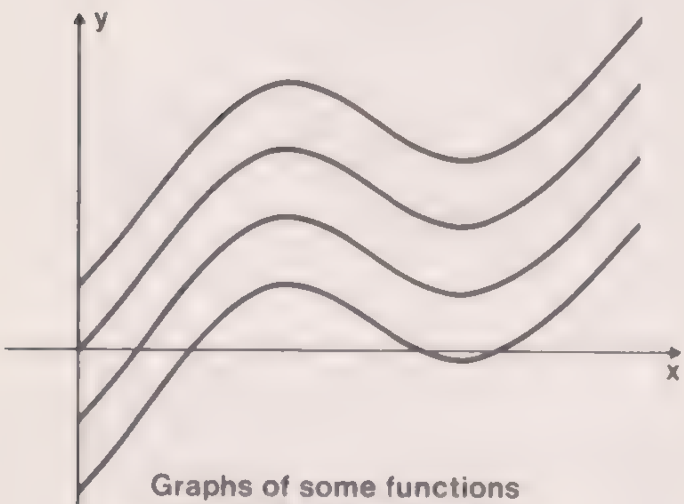
How can we use the result of the preceding section to evaluate integrals? To evaluate a definite integral involving a given function f, it is sufficient

to know a primitive function of f; the result in question helps us to recognize a possible primitive function, by telling us that every primitive function of a given continuous function f has the property that its derived function is f. Accordingly, if we look among the functions which have derived function f, we shall find all the primitives of f, but perhaps some other functions as well. Thus the result narrows the field in which to search for primitive functions of f, but it does not tell us how to be sure of finding them, or even how to be sure whether a supposed primitive function of f really is one or not.

In this section we shall demonstrate a further result which removes any doubt, by showing that under suitable conditions the "other functions" do not exist: every function with derived function f is in fact a primitive function of f. In other words we shall show that the shaded region in our diagram represents an empty set.

The principal step is to characterize the set in which the primitive functions of f are to be found: the set of all functions with derived function f. In terms of graphs, this set is the set of all functions whose graphs have slope $f(x)$ at each point with x-co-ordinate x. The following diagram shows, at the top, the graph of a continuous function f and, at the bottom, graphs of a few functions with derived function f.





Graphs of some functions with derived function f

This diagram indicates that the functions with derived function f have graphs that are congruent curves. That is, any one of the curves can be superimposed on any other by shifting it in a direction parallel to the y-axis: such a shift alters neither the x-co-ordinate of any point on the curve nor the slope, $f(x)$, at that point.

Such a shift in the graph is equivalent to adding a constant function to the original function; that is, replacing a function such as $x \longmapsto F(x)$ by $x \longmapsto F(x) + c$ where c is a real constant giving the amount of the shift. This demonstrates that all the functions with derived function f differ by constant functions. Like most arguments based on diagrams, this is a demonstration, not a proof.

So if F and F_1 both have derived function f, they differ only by a constant function. But we have seen in Exercise 9.1.3 that if F is a primitive of f, then any function that differs from F only by a constant function is also a primitive. Thus every function that has derived function f is a primitive of f, and so the shaded region of the diagram on page 229 represents an empty set.

The result that we have just demonstrated is the second part of the Fundamental Theorem of Calculus. It tells us that, to find some primitive function of a given continuous function f, it is sufficient to find any function whose derivative is f. A concise statement of the result can be obtained by denoting one of the functions with derived function f by F, so that $f = DF$; then the result tells us that F is a primitive of DF, or more precisely that

> If F is a real function whose domain includes the interval $[a, b]$, and if DF is continuous in $[a, b]$, then
>
> $$\int_a^b DF = F(b) - F(a).$$

Because expressions like $F(b) - F(a)$ occur frequently, we abbreviate by writing

$$[F]_a^b = F(b) - F(a)$$

so that, for example

$$[x \longmapsto x^2]_2^3 = 3^2 - 2^2 = 5.$$

To illustrate how this second part of the Fundamental Theorem of Calculus is used to evaluate integrals, let us apply it to

$$\int_1^2 x \longmapsto x^3.$$

We look for a function F such that

$$DF = x \longmapsto x^3.$$

We showed in Chapter 8 that differentiation always reduces the degree of a polynomial function by one, so we are led to consider $D(x \longmapsto x^4)$, which is $x \longmapsto 4x^3$. Apart from the factor 4, this is just what we want, and so a suitable function F is $x \longmapsto \frac{1}{4}x^4$. Hence, we have

$$\begin{aligned}\int_1^2 x \longmapsto x^3 &= \int_1^2 D(x \longmapsto \tfrac{1}{4}x^4)\\ &= [x \longmapsto \tfrac{1}{4}x^4]_1^2\\ &= \tfrac{1}{4}(2^4) - \tfrac{1}{4}(1^4)\\ &= \tfrac{15}{4}.\end{aligned}$$

The function $x \longmapsto x^3$ is continuous, so our application of the theorem is justified.

Exercise 1

Use the Fundamental Theorem of Calculus and standard derived functions, to evaluate

(i) $\int_0^t \exp$ (ii) $\int_0^\pi \cos$ (iii) $\int_0^{\pi/2} \sin$.

Could you have evaluated the second integral in a simpler way?

(HINT: Draw the graph of the cosine function and interpret the integral in terms of the graph; use the symmetry of the curve.)

The Fundamental Theorem of Calculus can be summarized by the statement (valid if f is a continuous function with codomain R and domain R or an interval of R) that F is a primitive function of f if and only if f is the derived function of F; this statement is equivalent to the pair of formulas

$$D\left(x \longmapsto \int_a^x f\right) = f$$

$$\int_a^b DF = F(b) - F(a).$$

The second of these formulas is particularly useful because it enables us to evaluate any integral if we can express the integrand (that is, the function to be integrated) as the derived function of another function.

9.4 Additional Exercises

Exercise 1

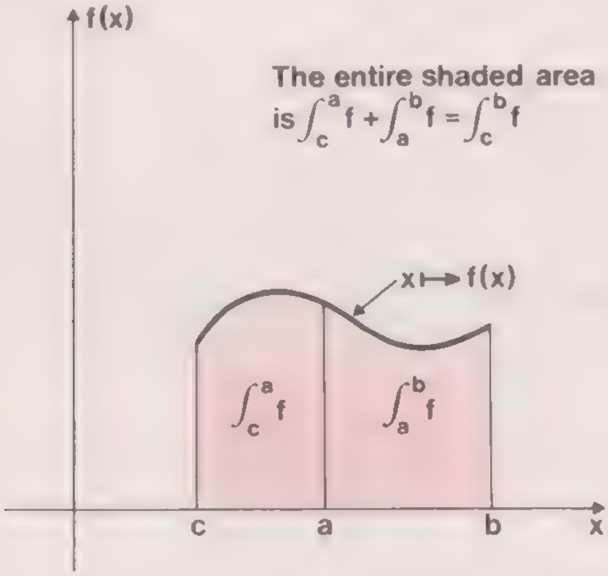
Use the expression

$$\int_c^a f + \int_a^b f = \int_c^b f$$

(see also diagram below) to show that, if f is a real continuous function with domain R, then for any real number c the function F given by

$$F : x \longmapsto \int_c^x f \qquad (x \in R)$$

is a primitive function of f.



Exercise 2

Find the derivative at x of each of the functions

$$x \longmapsto \int_a^x f \quad \text{and} \quad x \longmapsto \int_x^b f,$$

where f is a real function and a and b belong to the domain of f.

(HINT: Reconsider Exercise 1 above.)

Exercise 3

Which of the following are valid applications of the Fundamental Theorem of Calculus? If an application is not valid, explain why the theorem is not applicable.

(i) $\int_{-1}^{1} f_1 = [F_1]_{-1}^{1}$ where $F_1 : x \longmapsto \begin{cases} \frac{1}{2}x^2 \text{ if } x > 0 \\ 0 \text{ if } x = 0 \\ -\frac{1}{2}x^2 \text{ if } x < 0 \end{cases}$ $(x \in R)$

and $f_1 : x \longmapsto |x|$ $(x \in R)$.

(ii) $\int_{-1}^{1} f_2 = [f_1]_{-1}^{1}$ where $f_2 : x \longmapsto \begin{cases} 1 \text{ if } x \geqslant 0 \\ -1 \text{ if } x < 0 \end{cases}$ $(x \in R)$

and f_1 is given in (i).

(iii) $\int_{-1}^{1} f_3 = [f_2]_{-1}^{1}$ where $f_3 : x \longmapsto 0$ $(x \in R, \quad x \neq 0)$
and f_2 is given in (ii).

(iv) $\int_{-1}^{1} f_4 = [f_2]_{-1}^{1}$ where $f_4 : x \longmapsto 0$ $(x \in R)$
and f_2 is given in (ii).

9.5 Answers to Exercises

Section 9.1

Exercise 1

If F is a primitive function of f, then

$$\int_a^b f = F(b) - F(a)$$

for all a and b in the domain of f.

It follows that

$$\begin{aligned} \int_b^a f &= F(a) - F(b) \\ &= -(F(b) - F(a)), \end{aligned}$$

that is,

$$\int_b^a f = -\int_a^b f.$$

Exercise 2

f	$F(b)$	$F(a)$	F, a primitive function of f
$x \longmapsto 1 \quad (x \in R)$	b	a	$x \longmapsto x \quad (x \in R)$
$x \longmapsto x \quad (x \in R)$	$\frac{b^2}{2}$	$\frac{a^2}{2}$	$x \longmapsto \frac{1}{2}x^2 \quad (x \in R)$
$x \longmapsto x^2 \quad (x \in R)$	$\frac{b^3}{3}$	$\frac{a^3}{3}$	$x \longmapsto \frac{1}{3}x^3 \quad (x \in R)$

Exercise 3

Yes: *any* real number c gives a primitive function of f. To test whether F_c is a primitive function of f we must test whether

$$\int_a^b f = F_c(b) - F_c(a) \qquad ((a, b) \in R \times R).$$

Since F is a primitive function of f, we have:

$$\int_a^b f = F(b) - F(a) \qquad ((a, b) \in R \times R),$$

and therefore

$$\int_a^b f = (F(b) + c) - (F(a) + c) \qquad ((a, b) \in R \times R);$$

that is,

$$\int_a^b f = F_c(b) - F_c(a) \qquad ((a, b) \in R \times R),$$

so F_c is a primitive function of f.

Exercise 4

A primitive function of $x \longmapsto x$ is $x \longmapsto \frac{1}{2}x^2$, so that a more general primitive function of $x \longmapsto x$ is $x \longmapsto \frac{1}{2}x^2 + c$, where c is any real number. Denoting this function by F, we have

$$F(x) = \tfrac{1}{2}x^2 + c, \text{ and hence } F(0) = c.$$

The exercise requires $F(0) = 1$, so that $c = 1$, and therefore the required primitive function is

$$x \longmapsto \tfrac{1}{2}x^2 + 1 \qquad (x \in R).$$

Section 9.2

Exercise 1

(i) FALSE. The statement asserts that the cosine function is a primitive function of the sine function. Differentiating the primitive, cos, should restore the original function, sin, but in fact we have $D\cos = -\sin$, so the assertion given is false.

(ii) The statement asserts that sin is a primitive of cos; if this is so, then we should have $D\sin = \cos$, which is true; so there is no evidence against the assertion and we mark it TRUE. Notice the caution implied by our choice of words. We have argued in the previous text that*

$$(F \text{ is a primitive function of } f) \Rightarrow (DF = f),$$

but we have not shown that

$$(DF = f) \Rightarrow (F \text{ is a primitive function of } f)$$

which is the result we require here.

Section 9.3

Exercise 1

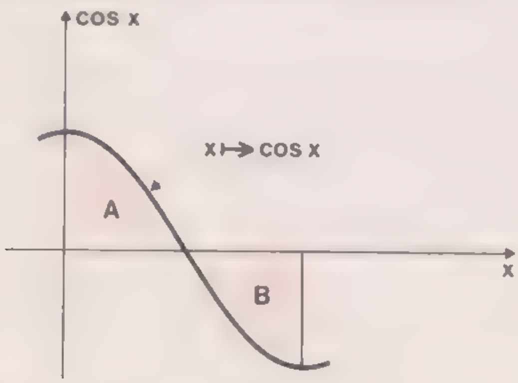
(i) $\displaystyle\int_0^t \exp = \int_0^t D\exp = [\exp]_0^t = \exp(t) - \exp(0) = e^t - 1$

(ii) $\displaystyle\int_0^\pi \cos = \int_0^\pi D\sin = [\sin]_0^\pi = \sin\pi - \sin 0 = 0 - 0 = 0$

(iii) $\displaystyle\int_0^{\pi/2} \sin = \int_0^{\pi/2} D(-\cos) = [-\cos]_0^{\pi/2} = -\cos\frac{\pi}{2} + \cos 0$

$$= -0 + 1 = 1.$$

An alternative method for the second case is:



* $\Rightarrow$ is the logic symbol for implication.

$$\int_0^{\pi} \cos = \int_0^{\pi/2} \cos + \int_{\pi/2}^{\pi} \cos$$
$$= \text{area } A - \text{area } B$$
$$= 0, \quad \text{by symmetry.}$$

The area B contributes negatively to the integral because the curve is below the x-axis. The total *area* is

$$\int_0^{\pi/2} \cos - \int_{\pi/2}^{\pi} \cos = [\sin]_0^{\pi/2} - [\sin]_{\pi/2}^{\pi} = 2.$$

Section 9.4

Exercise 1

The result $\int_c^a f + \int_a^b f = \int_c^b f$ gives

$$\int_a^b f = \int_c^b f - \int_c^a f$$
$$= F(b) - F(a) \qquad \text{by definition of } F,$$

and hence F is a primitive function of f. Notice that, since c is arbitrary, this method enables us to define as many different primitive functions of f as we please.

Exercise 2

Writing

$$F_1 \text{ for the function } x \longmapsto \int_a^x f$$

and

$$F_2 \text{ for } x \longmapsto \int_x^b f,$$

we see (by the result of Exercise 1 above) that F_1 is a primitive function of f, so that the Fundamental Theorem gives

$$DF_1 = f,$$

that is, the derivative of $x \longmapsto \int_a^x f$ at x is $f(x)$.

For F_2, we can use the result:

$$\int_x^b f = -\int_b^x f$$

obtained in Exercise 9.1.1; this gives

$$F_2 : x \longmapsto -\int_b^x f;$$

that is,

$$-F_2 : x \longmapsto \int_b^x f.$$

Using the result of Exercise 1 above again, we see that $-F_2$ is a primitive function of f, and so

$$D(-F_2) = f$$

By the First Rule of Differentiation

$$D(-1 \times F_2) = -1 \times DF_2,$$

and therefore

$$DF_2 = -f;$$

that is, the derivative of $x \longmapsto \int_x^b f$ at x is $-f(x)$.

These results give an alternative, very convenient formulation of the first part of the Fundamental Theorem.

Exercise 3

(i) Valid. f_1 is continuous everywhere in its domain, and F_1 is continuous and differentiable everywhere in its domain.
When $x < 0$, $|x| = -x$, and we have $DF_1(x) = -x$.
When $x \geqslant 0$, $|x| = x$, and we have $DF_1(x) = x$.
It follows that $DF_1 = f_1$.

(ii) Not valid (although the equation given is in fact true), because f_2 is not continuous (its graph has a gap at 0). To derive the given equation from the Fundamental Theorem, the integral must first be split into two parts: $\int_{-1}^{1} f_2 = \int_{-1}^{0} f_2 + \int_{0}^{1} f_2$, and the Fundamental Theorem applied to each part separately.

(iii) Not valid, because there is a gap in the domain of f_3. In fact we have $\int_{-1}^{1} f_3 = 0$, but $[f_2]_{-1}^{1} = 2$.

(iv) Not valid, because f_4 is not the derived function of f_2 (f_2 has no derivative at 0). Again, $\int_{-1}^{1} f_4 = 0$, but $[f_2]_{-1}^{1} = 2$.

Index

A reference in **bold type** indicates that a definition of the term appears on that page.

Index

Index

Index

AN INTRODUCTION TO CALCULUS AND ALGEBRA

Volume 1 Background to Calculus

This first volume provides the background from which calculus can be developed. It begins with the concept of *set*, and goes on to introduce *mappings*, *functions*, *sequences* and *series*, and to discuss *limits* and *convergence* starting from an intuitive standpoint. Making use of these concepts, the basic idea of *integration* and *differentiation* are then presented, and the final chapter links the two by means of the *fundamental theorem of calculus*.

This is the first of three volumes presenting some of the essential concepts of mathematics, a few important proofs (usually in outline), together with exercises designed to reinforce the understanding of the concepts and develop the beginnings of technical skill.

A particular feature of these volumes is the use of two-colour printing to emphasize important concepts and definitions and to heighten the impact of diagrams. Exercises are provided throughout, and detailed answers (with additional comment) are given at the end of each chapter.

The selection of material has been made with the needs of students of other subjects particularly in mind.

The material presented is, however, suited to a wide class of reader and will provide both a modern introduction and a handy reference to two important areas of mathematics: Calculus and Algebra.

Volume 1 Background to Calculus
Volume 2 Calculus Applied
Volume 3 Algebra

The Open University Press

THE OPEN UNIVERSITY PRESS

SBN 335 00002 9